Brain-based Authentication Systems and Brain Liveness Problem

by

Mohammad Javad Sohankar Esfahani

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved July 2021 by the
Graduate Supervisory Committee:

Sandeep K. S. Gupta, Chair
Ayan Banerjee
Partha Dasgupta
Marco Santello

ARIZONA STATE UNIVERSITY

August 2021

ABSTRACT

In recent years, brain signals have gained attention as a potential trait for biometric-based security systems and laboratory systems have been designed. A real-world brain-based security system requires to be usable, accurate and robust. While there have been developments in these aspects, there are still challenges to be met. With regard to usability, users need to provide lengthy amount of data compared to other traits such as fingerprint and face to get authenticated. Furthermore, in the majority of works, medical sensors are used which are more accurate compared to commercial ones but have a tedious setup process and are not mobile. Performance wise, the current state-of-art can provide acceptable accuracy on a small pool of users data collected in few sessions close to each other but still falls behind on a large pool of subjects over a longer time period. Finally, a brain security system should be robust against presentation attacks to prevent adversaries from gaining access to the system.

This dissertation proposes E-BIAS (EEG-based Identification and Authentication System), a brain-mobile security system that makes contributions in three directions. First, it provides high performance on signals with shorter lengths collected by commercial sensors and processed with lightweight models to meet the computation/energy capacity of mobile devices. Second, to evaluate the system's robustness a novel presentation attack was designed which challenged the literature's presumption of intrinsic liveness property for brain signals. Third, to bridge the gap, I formulated and studied the brain liveness problem and proposed two solution approaches (model-aware & model agnostic) to ensure liveness and enhance robustness against presentation attacks. Under each of the two solution approaches, several methods were suggested and evaluated against both synthetic and manipulative classes of attacks (a total of 43 different attack vectors). Methods in both model-aware and model-agnostic approaches were successful in achieving an error rate of zero (0%).

More importantly, such error rates were reached in face of unseen attacks which provides evidence of the generalization potentials of the proposed solution approaches and methods. I suggested an adversarial workflow to facilitate attack and defense cycles to allow for enhanced generalization capacity for domains in which the decision-making process is non-deterministic such as cyber-physical systems (e.g. biometric/medical monitoring, autonomous machines, etc.). I utilized this workflow for the brain liveness problem and was able to iteratively improve the performance of both the designed attacks and the proposed liveness detection methods.

ACKNOWLEDGMENTS

Whoever does not thank people indeed hasn't thanked god.

(Islamic Theology)

Human beings are members of a whole
In creation of one essence and soul
If one member is afflicted with pain
Other members uneasy will remain
If you have no sympathy for human pain
The name of human you cannot retain
(Saadi, Persian poet (1210-92))

I wanted to acknowledge that me writing this sentences is possible due to dreams, sufferings, resistance, emotions, senses, feelings, pains, revolts, and intellectual power of all and each of the humanity from the dawn of consciousness to today. I stand here on the shoulders of all oppress(ed)(or)s and love(d)(r)s whom shaped the history and build the civilizations we seem proud of them and do need to acknowledge that I am no better than any other, and also one of those who has been oppressor, oppressed, loved and lover.

This manuscript is the outcome of my family's support, my advisors' and committee guidance, my colleagues' collaboration, my friends encouragements and of course my faults.

Without the love of my mom, Maryam, and my dad, Ahmad, kindness of my sister, Fateme, and my brother, Hadi, there wouldn't be any Javad to brag about a PhD thesis. I love each of you individually and all of you collectively and please forgive my endless shortcomings.

My advisor, Dr. Sandeep Gupta, teached me concepts beyond scientific research and I am always grateful for his time, thoughts, trust and allowing me to explore and learn from my failures. Thanks Sandeep.

My co-advisor, Dr. Ayan Banerjee had a significant role in me learning how to conduct research, never losing hope, working and working hard, going out of the box and being kind to everyone regardless of their past. Thanks ayan.

Thanks to all my colleagues and labmates which we laughed and panicked together and made these years fly by. I owe much to Koosha, who introduced me to brain research, closely collaborated with me, co-authored numerous works, made so many great memories and of course helped my like a true friend. Thanks Koosha.

I could not have accomplished much if it was not for numerous friends in different parts of the world, in Phoenix, US, Canada, Europe and my beloved country, Iran. Among them, I will just mentioned two names as I spent a great deal of time with them while impersonating as a researcher. Asa and Bahman which with their endless kindness, support, love and sense of humor empowered me to finish my degree and I will always be there for you. I am confident future belongs to you and do not forget that sky is the limit.

I have one final acknowledgment to every and each of the great thinkers and scholars of human thought history, and as this work is concerned with brain -symbol of consciousness- and its liveness, I end this section with the introduction of the book ”Žižek’s Jokes: (Did You Hear the One about Hegel and Negation?)” by philosopher Slavoj Žižek, to remind us to not take ourselves that serious.

"The Role of Jokes in the Becoming-Man of the Ape

One of the popular myths of the late Communist regimes in Eastern Europe was that there was a department of the secret police whose function was (not to collect, but) to invent and put in circulation political jokes against the regime and its representatives, as they were aware of jokes' positive stabilizing function (political jokes offer to ordinary people an easy and tolerable way to blow off steam, easing their frustrations). Attractive as it is, this myth ignores a rarely mentioned but nonetheless crucial feature of jokes: they never seem to have an author, as if the question "who is the author of this joke?" were an impossible one. Jokes are originally "told," they are always-already "heard" (recall the proverbial "Did you hear that joke about . . . ?"). Therein resides their mystery: they are idiosyncratic, they stand for the unique creativity of language, but are nonetheless "collective," anonymous, authorless, all of a sudden here out of nowhere. The idea that there has to be an author of a joke is properly paranoiac: it means that there has to be an "Other of the Other," of the anonymous symbolic order, as if the very unfathomable contingent generative power of language has to be personalized, located into an agent who controls it and secretly pulls the strings. This is why, from the theological perspective, God is the ultimate jokester. This is the thesis of Isaac Asimov's charming short story "Jokester," about a group of historians of language who, in order to support the hypothesis that God created man out of apes by telling them a joke (he told apes who, up to that moment, were merely exchanging animal signs, the first joke that gave birth to spirit), try to reconstruct this joke, the "mother of all jokes." (Incidentally, for a member of the Judeo-Christian tradition, this work is superfluous, since we all know what this joke was: "Do not eat from the tree of knowledge!" — the first prohibition that clearly is a joke, a perplexing temptation whose point is not clear."

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

Table Page

Table Page

LIST OF FIGURES

Figure Page

Chapter 1

INTRODUCTION AND BACKGROUND

Human life has become more than ever dependent on Cyber Physical Systems (CPS) in nearly all domains (e.g. medical, transportation, communication, military, etc.), and therefore more than ever target of attacks (Khaitan and McCalley, 2015). These systems, control and affect numerous aspects of individual and social life (e.g. safety, security, financial, identity, reputation, etc.), and therefore their security and robustness is of critical importance. Cyber Physical Systems (CPS) are being increasingly deployed due to two reasons. First, advances in data driven modeling (either grounded in mathematical models or statistical machine learning models) that enhances performance of such systems. Second, rapid growth of Internet of Things (IoT) infrastructure and mobile devices which allows pervasive data gathering (through sensors). On one hand, these reasons have caused CPS to be highly used in numerous domains and by large number of general users. On the other hand, this wide-spread and pervasive usage has made them interesting targets for adversaries (Humayed *et al.*, 2017; Yampolskiy *et al.*, 2013; Taylor and Sharif, 2017).

One of the major types of security breaches in CPS is called *Presentation Attacks* (PA) where an adversary presents malicious data to a system in order to alter its behavior (Ramachandra and Busch, 2017; Sousedik and Busch, 2014; Raghavendra and Busch, 2014). This type of attack also includes attempts to impersonate valid users (by presenting old/fake data) which then grants complete control of the system. To prevent such attacks, systems are equipped with access control mechanisms (Abdulkader *et al.*, 2015) which are of three kind in nature; 1) *Knowledge Based*: something you know e.g. a password, 2) *Token Based*: something you have e.g. a token/card, 3)

Biometrics Based: something you are, which means human traits such as fingerprint, face, etc. The third mechanism (i.e. use of human traits) also known as *biometric authentication* is the current trend in diverse range of applications from military facilities to airport security checks to smart phones (Wayman *et al.*, 2005).

Currently, CPS systems are more than ever using biometric-based access control mechanisms due to its advantages such as being unique to the user, always carried by user, and ease of use (Jain *et al.*, 2016). However, three security challenges also exist; First, traits can be collected/stolen easily without user knowledge (e.g. fingerprints on object/surface, face images on Internet or captured by camera, voice being recorded, etc.). Second, traits can be artificially generated with relatively low effort (e.g. gummy finger, face model and mask, generating heart/brain signal) (Nixon *et al.*, 2008; Xu *et al.*, 2016; Sadeghi *et al.*, 2017). Finally, there is no deterministic method to check an input trait against the stored signature of the trait in system. While traditional inputs (e.g. password) can be checked deterministically, human traits are checked using non-deterministic methods (e.g. probabilistic, approximation, heuristic, etc) such as Machine Learning techniques. Deterministic methods are error-free while non-deterministic methods inherently have errors which is the root cause of *Presentation Attacks* (PA). This inherent error allows for adversarial input to be detected as valid input which in turns provides access to system for the adversary.

The three above-mentioned reasons makes biometric authentication systems vulnerable to PA where an adversary presents old (i.e. previously collected) or fake (i.e artificially generated) traits to access the system. This fundamental vulnerability forces a reconsideration of the current high levels of trust and confidence in such systems. This level of trust has originated from two assumptions; 1) human traits are unique to each individual, 2) traits are almost impossible to be mimicked. In recent years, it has been shown that exploiting biometric systems by mimicking human traits (i.e. PA) can be done fairly easy (Nixon *et al.*, 2008; Xu *et al.*, 2016; Sadeghi *et al.*, 2016b). This opens door to disastrous problems since biometrics are generally considered to be the most secure and robust of access control mechanisms and hence are used in critical systems. Therefore, to prevent such events there is an immediate need for developing methods to prevent/detect PA to improve the trust level and security of theses systems. In literature these methods are termed as *Biometric Liveness Detection (BLD)* (Akhtar *et al.*, 2015; Matthew and Anderson, 2014b; Matthew, 2016; Rogmann and Krieg, 2015) or *Presentation Attack Detection (PAD)* (Ramachandra and Busch, 2017; Sousedik and Busch, 2014; Raghavendra and Busch, 2014). In this manuscript we will be using the *Biometric Liveness Detection* (BLD) term, or simply Liveness Detection (LD).

Lets consider a futuristic scenario which current technology is not much far from it; Imagine relaxing in your self-driving car after a long day at work, checking your banking application on your smartphone equipped with fingerprint sensor while the artificial pancreas in your body is monitoring blood glucose level and pumping insulin. Upon arriving, your smart-home recognizes the car and opens the garage door. *Alan Turing*, your household robot's nickname, greets you with a hot coffee and asks what dish you would like to be cooked for dinner. You head to bedroom -your most private/personal room- which at the door recognizes your face and voice before opening. In the described scenario, human traits control all the mentioned systems and an adversary capable of mimicking the traits can launch a series of PA in order to change self-driving car destination, access your phone and bank accounts, alter the insulin release rate of the artificial pancreas, enter your home and command your household robot, and invade your bedroom privacy. One might say the described situation is distant from current state art of in technology and artificial intelligence (AI), but we should keep in mind that there is already self-driving cars on the road. This implies that we are convinced and confident about the capabilities of machines in fields such as AI, computer vision, and image/signal processing that we allow machines to perform safety-critical tasks such as driving cars and airplanes. In the same manner, these capabilities can be used by adversaries to compromise machines and therefore the task of distinguishing between human and machine is of significant importance for future Cyber Physical systems (CPS).

Even today, one would interact and rely on numerous systems (which receive human traits) throughout a day for different tasks, many of them safety critical. These systems are typically designed to only provide access and accept commands from valid users and reject others. Distinguishing between users is done by Access Control Module (ACM) which can have two type of errors; rejecting valid user (*False Reject*) or accepting invalid user (*False Accept*). While false reject error reduces usability for a valid user and causes frustration, False Accept (FA) error (which PA is based on it) grants system control to an invalid user (i.e. adversary) which poses serious threats. It is noteworthy that if a system is not equipped with ACM, then it will accept commands from any entity, opening door to major security vulnerability. Example of such systems is virtual personal assistants in smart-phones (e.g. Apple Siri) and smart-speakers (e.g. Amazon Echo) which accept voice commands from any human/machine capable of generating/replaying voice commands. Beside that some cars, household embedded systems (e.g. refrigerator, oven, etc.) and smart home devices (e.g. door, window, A/C, audio-video systems, etc.) also operate based on human inputs such as voice. Another important category is medical systems that monitor and operate on human body (e.g artificial pancreas, heart) purely based on human traits.

Lets consider another scenario; there is a set of robots/drones that receive voice commands from human users. The problem here is again how to prevent PA which requires ensuring that the set of machines only accept inputs from valid human user and not another machine? Two cases to be considered are: 1) Machines are equipped with Access Control Module (ACM): A possible attack can happen if one of these machines is exploited (or an outsider adversary machine) records the commands of a valid user and later on replays it to control the other machines. A more complex attack can be launched if the attacking machine uses machine learning algorithms to learn and mimic valid user voice which in turn will enable it to generate any desired command and not only replay the previously recorded ones. 2) Machines are not equipped with ACM: as discussed earlier in the case of virtual personal assistants and smart-speakers, any human or machine can easily attack and control these machines. It is noteworthy that the described scenario can cause serious threats in case of say military drones.

The two scenarios share a common problem; *how to prevent presentation attacks which is dependent on detection of adversarial inputs.* In the first described scenario, human provides adversarial inputs while in the second scenario machine does. Adversarial inputs can be defined as an input from human/machine adversary which mimics valid user trait and it can be categorized into three types; 1) Exact replay of a legitimate trait, 2) Altered version of a legitimate trait, 3) Artificially generated traits. With high confidence it can be said that, biometric systems will accept the first and second type of inputs, and for the third type, acceptance rate depends on how well it has been crafted similar to the genuine trait. Biometric authentication system by nature is incapable of distinguishing between these forged inputs and genuine ones, since it is only designed to match the input trait with the stored ones and nothing more. Therefore improving the performance of the authenticator will not solve the issue and a different set of approaches is required to deal with the problem, which directs us to biometric liveness detection (BLD) methods.

This thesis focuses on authentication and liveness problem in systems which use brain signal as their input trait. Brain signals are becoming a potential human trait for authentication tasks in real-world systems due to two reasons. First, brain signals contain information which are unique to an individual, nearly impossible to impersonate without invading personal space, and chaotic over time. This is markedly different from biometrics such as fingerprints, voice, and face, which can be captured without the subject's knowledge or purposefully altered (Hu *et al.*, 2011). Moreover, seamless availability of EEG data opens up potential usage in securing personal information in scenarios where a password may not be entered, spoken out, or remembered. For example, the notion of "hands-free" security can be imagined, when the person is driving or pre-occupied with other tasks and cannot focus on targeted security related tasks (Banerjee *et al.*, 2013). Furthermore, humans have functioning brain until the very last stages of life while other traits can be lost during lifetime. So for people with disabilities (e.g. Amputation, Blindness, Muteness) brain-based authentication can be used instead of other biometric authentication systems. In some environments where its dark/noisy or subject has gloves on brain-based authentication can perform well while systems such as face recognition, voice recognition and fingerprint will face serious challenges. Another use case would be seamless authentication for Augmented/Virtual Reality systems. Second, availability of wireless, easy-to-wear, non-invasive and low cost sensors that can capture EEG signals (i.e. Electroencephalography) from human scalp and send them to other devices. However, brain-based authentication systems are still in research phase and there is no commercial brain authentication system.

In biometrics literature, there are numerous works which provide different approaches for solving liveness detection problem in case of traits such as fingerprint, face, iris and voice (for surveys see Ramachandra and Busch, 2017; Marasco and Ross, 2015; Czajka and Bowyer, 2018; Wu *et al.*, 2015a). However, there are no works on liveness detection for brain signals since it is commonly assumed that these signals possess an intrinsic liveness property (Zhao *et al.*, 2019; Kong *et al.*, 2018; Maiorana and Campisi, 2017; Garau *et al.*, 2016; Thomas and Vinod, 2016; Sundararajan *et al.*, 2015; Fraschini *et al.*, 2014; Nakanishi *et al.*, 2009). Furthermore, as an solution approach it is suggested that the vulnerable Biometric systems (e.g. fingerprint, face, iris and voice) be augmented with a secondary input of type brain signals to ensure liveness detection. In this thesis 4.2, I proposed a novel presentation attack using artificial signals which successfully bypassed brain-based authentication systems, and challenged the assumption of intrinsic liveness property for brain signals. Therefore, brain waves similar to other traits are vulnerable to presentation attacks and their liveness is not guaranteed.

Brain liveness problem has not been recognized in the literature and in order to bridge the gap, this dissertation for the first time formulates the problem statement and systematically studies brain liveness problem. Two solution approaches (model-aware & model agnostic) are proposed and evaluated against two class of attacks (synthetic and manipulative) with 44 attack types. For each approach, I studied several methods and evaluated their performance against the attack dataset. For model-aware approach, which only EEG signals (not any attack samples) was used in tuning decision making parameter, error rates were less than 1%. In case of model-agnostic approach, under *normal* protocol (training set contains samples from each attack categories, but not all attack types), we achieved error rate of 0% , and in case of the more challenging *unseen* protocol (some attack categories are not included in training set) error rate was less than 1%. The low error rate (<1%) of model-aware and model-agnostic (under unseen protocol) solutions shows these approach do generalize well against new attack types, which is an significantly important feature of defense mechanism. Reason being that the attack space is extremely large (cannot be brute forced) and regardless of the number of attacks utilized during system design new ones can and will emerge.

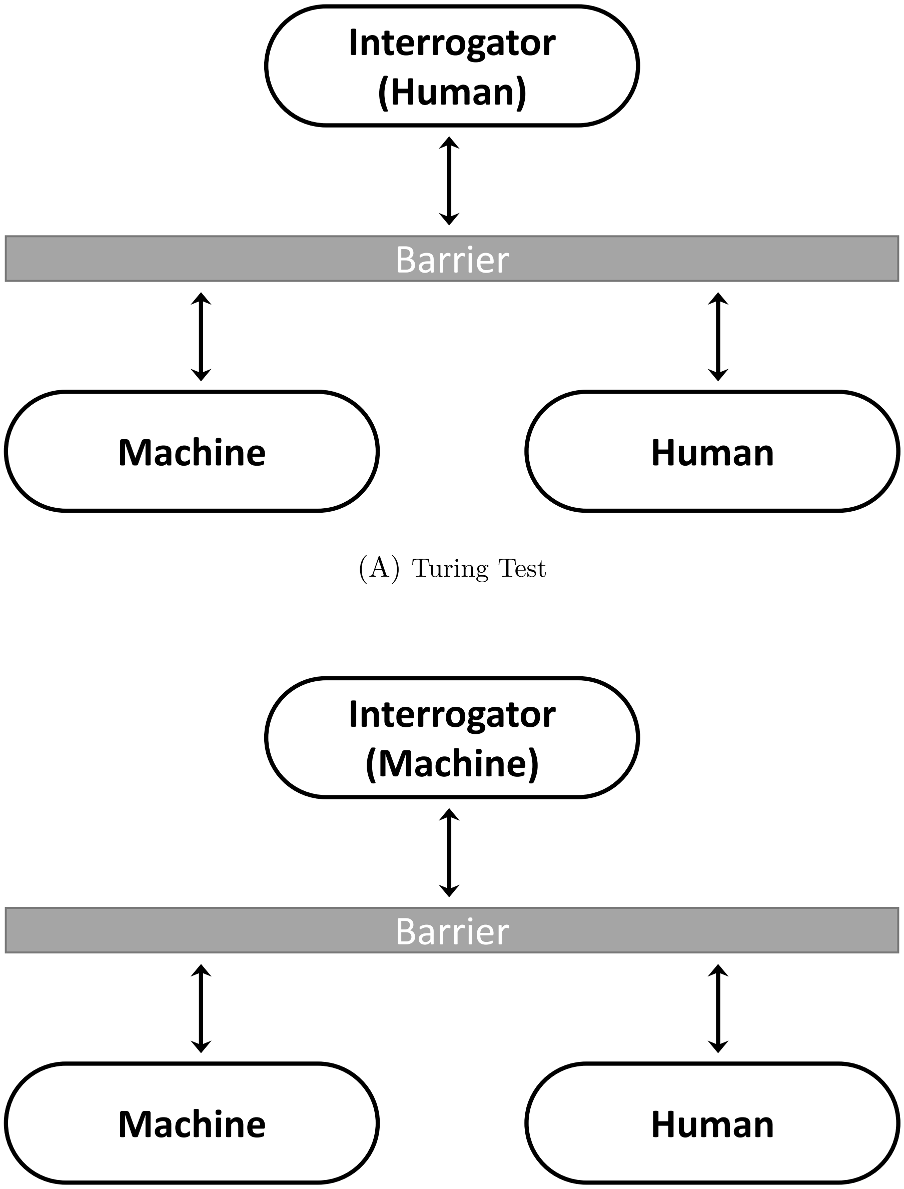
Furthermore, in section 2.4, I suggest an adversarial work flow to facilitate liveness attack and defense cycle for domains in which decision-making process is non-deterministic such as cyber-physical systems (e.g. biometric/medical monitoring, autonomous machines, etc.). The work flow is aimed for perspective of both attacker and defender to aid attacker in crafting more effective attacks and help defender to design more robust liveness detection methods.

The rest of this manuscript is structured as follows; I make analogies between liveness detection with *Turing test* for artificial intelligence in rest of chapter one. In chapter two, I discuss the brain liveness problem in details and provide problem statement, definitions, system model and threat model. The two solution approaches are explained and their performance is evaluated against attack set. Chapter three, systematically studies biometric liveness detection problem, provides taxonomy of the state-of-the-art and suggest new approaches for biometric liveness. Chapter four, discusses my proposed system *E-bias* (EEG-based Authentication and Identification System), and explains a novel presentation attack using predictive models in feature-domain instead of time-domain. Finally chapter five concludes and discusses future research directions.

1.0.1 Liveness vs Turing Test

The ability to discern between human and machine is a new research question while in the dawn of Artificial Intelligence (AI) all the effort was on building machines as close as possible to the humans so that a human observer is incapable of discerning between the two. Seven decades back in year 1950, Alan Turing designed his well-know test, *imitation game*, as a way of measuring the intelligence level of a machine (Turing, 1950). The test works as follows; a human interrogator is interacting via only text with two (hidden) entities A and B, one being human and the other one machine, through written text via a monitor. Interrogator interacts with the two entities A and B for certain amount of time through asking questions and receiving their replies. Afterwards, if the interrogator can not differentiate between the human and the machine from their answers, than the machine has successfully passed the test. Turing suggests that if a machine passes this test, it can be concluded that it is as intelligent as the human. This test was designed in the Artificial Intelligence (AI) domain where the goal is to design smarter machines, while in security domain smarter machines are the problem and liveness tests are designed to make sure a machine can differentiate between a human and a machine.

In the liveness test, scholars are trying to automate a modified version of Turing test where human interrogator is replaced with a machine and the objective goal is negated as shown in Figure 1.1. In Turing test, the goal is to design a machine that can not be differentiated from a human, while in liveness test the goal is to design a machine that is capable of distinguishing between a machine and a human. The state-of-art in liveness tests mainly focuses on biology, while Turing test focused on intelligence, and that is the type of change needed for liveness test. In Turing test the difference in intelligence level between the humans and the machines was leveraged, while in the Liveness test scholars make use of the point that human body is an organic structure while machine is made of silicon. Taking into account the advances of Biofrabrication (i.e. organ manufacturing), biology based liveness test have limited time frame in front of them, and the shift toward intelligent checking and focus on brain capabilities of human which still machines haven't achieved should start sooner rather than later.



(A) Turing Test

(B) Liveness Test

Figure 1.1: Turing vs. Liveness Test.

Chapter 2

BRAIN LIVENESS PROBLEM

In this chapter, I will try to bridge the gap in literature with regard to brain liveness problem by discussing it in particular and also look into biometric liveness detection in general as needed (for a more detailed discussion on biometric liveness detection check chapter 3). This chapter will first provide an introduction on the topic and afterwards look into the problem statement, challenges, system and threat models. Afterwards, I discuss the proposed adversarial cycle and solutions approaches for brain liveness problem, and their performance evaluation. It is noteworthy that the adversarial cycle is not limited to brain liveness problem, and is effective for systems where inputs from physical world are processes in non-deterministic way such as cyber-physical systems like biometric systems, autonomous vehicles and smart homes.

2.1 introduction

Universality and permanency of the human brain has resulted in the prospective use of brain signals in several domains including biometric security. The advent of wearable and implanted brain sensors (e.g. Neuralink) and the decade long US Brain Intuitive (2014-25) with more than $4.5 billion in funding are just samples of such a trend. The rationale of using brain in security has always been centered around its inherent inaccessibility (remote sensing is not possible), and the high entropy of the signals that can be measured such as electroencephalogram (EEG), fMRI, Petscan. Usage of brain also enables hands-free cyber-physical security systems for users preoccupied with another task(s). While in cryptography, security guarantee is based on randomness of the key and backed by mathematical theories, in biometric systems robustness against spoofing attacks depends on nature of the input in use (fingerprint, face, voice, brain). In biometric field, there is an consensus that brain signals are the ideal option due to the chaotic nature of the measured signals (e.g. EEG). This assumption is often referred to as the *intrinsic liveness property* of brain and is primarily backed by the point that electroencephalogram (EEG) signals are outcome of numerous neuron activities which get affected by surrounding contexts and past experiences (Zhao *et al.*, 2019; Kong *et al.*, 2018; Maiorana and Campisi, 2017; Garau *et al.*, 2016; Thomas and Vinod, 2016; Sundararajan *et al.*, 2015; Fraschini *et al.*, 2014; Nakanishi *et al.*, 2009). However, there is a lack of quantitative evidence on the assumption of intrinsic liveness and brain signal being an ideal source of randomness and entropy.

Beside inherent entropy levels, robustness in security systems would depend on how well the current state-of-the-art in feature extraction and modeling techniques are capable of utilizing the full potential of the available randomness in brain signal. If the brain signal randomness is low in the first place or the approach in which brain signal is processed for decision making significantly lowers the effective randomness in use, then adversaries would be capable of crafting inputs which can mimic brain signal behavior. In section 4.2 and current chapter, I provide manifestations of predictive models that can artificially generate brain signals (EEG) which can spoof biometric authentication systems that use EEG as their primary signal source. This casts doubts on the consensus of presuming intrinsic liveness for brain signal and their inherent robustness against attacks. To shed light on the topic, I study to what extent brain signals are suitable for guaranteeing *liveness* of the entity which is interacting with biometric system. Furthermore, we investigated which categories of attacks can be launched and what are the counterattacks. I created an array of attack vectors in two category of *manipulative* and *synthetic* attacks and tested their success in spoofing the biometric system. Afterwards, for counterattack, I proposed two solution approaches, *model-aware* and *model-agnostic*. For each approach, I studied several methods and evaluated their performance against our attack dataset. For model-aware approach, which only EEG signals (not any of the attack samples) were used in tuning decision making parameter, error rates were less than 1%. In case of model-agnostic approach, under *normal* protocol (training set contains samples from each attack categories, but not all attack types), we achieved error rate of 0% , and in case of the more challenging *unseen* protocol (some attack categories are not included in training set) error rate was less than 1%. The low error rate (<1%) of model-aware and model-agnostic (under unseen protocol) solutions shows these approach do generalize well against new attack types, which is an significantly important feature for a defense mechanism. Reason being that the attack space is extremely large (cannot be brute forced) and regardless of the number of attacks utilized during system design new ones can and will emerge.

In rest of introduction section, I briefly discuss authentication and liveness detection for human inputs and afterwards closely examine literature's reasons for brain signals intrinsic liveness property. I argue against those reasons and provide examples of presentation attacks that did spoof brain biometric systems. I introduce brain liveness detection problem, provide definitions, system model and problem statement, and then discuss advantages of brain signals and challenges for its liveness detection. In the related work section, we look into liveness detection methods for other human traits (face, fingerprint, iris, and voice) and compare them based on type, approach, and dataset for attack, size of subjects pool, and their performance (error rates). In the succeeding sections, I discuss the adversarial cycle, solutions approaches, threat model and experimental setup in details, and afterwards evaluate performance of the proposed solution, and finally provide future research directions and conclude.

2.1.1 Security of Human Input

Sending voice commands to smartphones, logging to laptop with face recognition, checking into workplace with fingerprint, conducting remote meeting/interview/exam using through users' video/voice, and controlling prosthetic body organs (e.g. arm or leg) with brain signals are few of the systems which depend on human inputs. With the fast growing trend of moving from traditional inputs (such as mouse, keyboard or even touchscreen) to using human inputs (e.g. fingerprint, face, voice, iris, psychological signals (heart and brain)), comes both improved user experience and usability and also new security and privacy challenges. The primary security challenges of such systems is to ensure two guarantees: first, input belongs to the claimed user, and second it has genuinely originated from live human at the current point in time. *Presentation Attacks* (PA) aim to bypass these two security guarantees, and allow adversaries to interact with system in place of the legitimate user (or at least cause the system to malfunction). In Presentation Attacks (PA) (also termed as *spoofing attack*), adversary provides either past genuine user inputs or fake/crafted inputs to get access to system. Such attacks have been successful in spoofing real world systems and they manifest in forms such as gummy finger, face masks, printed face/iris image, recorded/synthesized voice and manipulated/synthesized heart and brain signals.

To counterattack Presentation Attacks (PA), both *authentication* and *liveness detection* methods are required. Authentication checks if the input matches the claimed user, and liveness detection determines if the input is being sensed from a live human at the current point in time, and is not an past or crafted input. However, systems generally focus more on authentication mechanism and pay less attention to liveness detection. It is noteworthy that these two mechanisms provide mutually exclusive security guarantees, and one can not replace the other. In other words, if an incoming input matches user's signature it does not imply it indeed has originated from a live human, and vice versa.

Livneness Detection (LD) aims to ensure two properties in an input; *live* and *timely*. *Live property*: input has originated from a live human, and is not a manipulated or artificially generated input, and *Timely property* input belongs to current point in time, and does not belong to past. In high level, liveness goal is to ensure that the entity interacting with system is an live human being which is providing genuine inputs. On the Internet, CAPTCHA checks for liveness of the entity behind traditional inputs (mouse and keyboard), and similarly there is a need for liveness detection in case of human inputs.

2.1.2 Intrinsic Liveness Property

In the biometrics field, there exist a large body of work on liveness detection for human inputs. There exists two general direction: first, developing methods for ensuring liveness of the trait in use, and second using/adding a trait with intrinsic liveness property. The first direction, attempts to ensure liveness through detecting novel features in the trait (software-based), sensing new modalities/traits (hardware-based), presenting stimuli and detecting its expected impact (context-based) (check anti-spoofing handbook (Marcel *et al.*, 2019), or surveys on LD for face recognition (Ramachandra and Busch, 2017), fingerprint (Marasco and Ross, 2015), iris (Czajka and Bowyer, 2018), and voice recognition (Wu *et al.*, 2015a)).

In the second direction, liveness detection happens through using human traits which possess *'intrinsic liveness property'*, meaning they cannot be synthesized/mimicked and they are inherently robust against presentation (spoofing) attacks. Hence, such traits would automatically ensure liveness, and there would not be need for designing liveness detection methods for them. In literature, physiological signals and especially brain signals with its high entropy and chaotic nature are presumed to have intrinsic liveness property and would not require LD (Zhao *et al.*, 2019; Kong *et al.*, 2018; Maiorana and Campisi, 2017; Garau *et al.*, 2016; Thomas and Vinod, 2016; Sundararajan *et al.*, 2015; Fraschini *et al.*, 2014; Nakanishi *et al.*, 2009). On the contrary, there has been handful of works (one of them by me (Sadeghi *et al.*, 2017) discussed in section 4.2) that show both artificially generated signals (using hill climbing (Maiorana *et al.*, 2013) or predictive models (Sadeghi *et al.*, 2017)), and manipulated signals (by noise addition (Gui *et al.*, 2016; Sadeghi *et al.*, 2017)) can bypass brain-based authentication mechanisms that use machine learning models. To cast light on the discussion, I will look into the reasons provided by literature to back the brain intrinsic liveness claim.

2.1.3 Examining Brain Intrinsic Liveness Claim

The inconsistency, pointed out in the preceding section(2.1.2), between the presumed claim of brain intrinsic liveness supported by most of the literature and the challenging results of few works, asks for revisiting the reasons behind this claim. Zhao *et al.* (2019) states "... EEG has emerged as a good candidate for individual identification because of its advantages such as universality, intrinsic liveness detection capability, and robustness against attacks (Campisi and La Rocca, 2014)." (reference is updated based on this manuscript). In the cited reference, Campisi and La Rocca (2014) points to three factors; First, "brain signals are result of ionic current flows within the neurons of the brain in response to a specific task or during a specific mental state", second they cannot be captured at a distance and third they cannot be left on objects (unlike fingerprints). Based on these, authors conclude brain signals are "less likely to be synthetically generated" and the liveness problem which exist with other traits is "naturally overcome without the need to resort to specifically designed sensors". Maiorana and Campisi (2017), also focuses on impracticality of capturing brain signals from a distance and states "... in addition to the obvious universality and intrinsic liveness properties, they are also highly robust against presentation attacks, being their acquisition at a distance impossible at the present stage of technology.". Kong *et al.* (2018), mentions brain signals have "high concealment, non-stealing, and liveness detection" and "unforgeability" advantage which makes them a suitable option for systems with "high confidentiality and high-safety requirements". The only related reference authors provide is to another work by Maiorana *et al.* Maiorana *et al.* (2016)which points back to the already discussed Campisi and La Rocca (2014).

In other works, authors state brain signals have "robustness to spoofs and intrinsic liveness detection" (Garau *et al.*, 2016), "robustness against spoofing attacks, ..., intrinsic liveness detection" (Thomas and Vinod, 2016), "... it is almost never possible to circumvent an EEG because spoofing a brainwave is not possible. Liveness detection is inherent in the measurement phase ..." Sundararajan *et al.* (2015) (2015), " robustness against spoofing, ... and liveness detection" (Fraschini *et al.*, 2014), "to be robust against spoofing attacks to the sensor being a signal that cannot be observed and that therefore cannot be synthesized by an attacker." (Campisi *et al.*, 2011), "generated by the activities of neurons in a brain cortex ... it is effective for anti-circumvention. Of course, the brain wave possesses the function of liveness detection since it is generated by only live human beings." (Nakanishi *et al.*, 2009).

In summary, there seems to be two factors behind the claim of brain intrinsic liveness. First, brain signals are outcome of numerous neurons in response to some specific task or mental activity and therefore an adversary cannot synthesize such a signal. However, the catch is that the adversary does not need to craft the exact signal which an authentic brain would produce, but instead an input around that range would be sufficient. In other words, issue is not with the brain as a source of input, but instead with how we process these inputs. Brain signals compared to other human inputs have higher entropy, a chaotic nature and a better source of randomness but the processing and decision making procedure that happens on them allows for attacks to happen. Authentication and liveness detection goals are accomplished through nondeterministic approaches (e.g. pattern recognition and machine learning) since as of now there are no deterministic methods as with checking passwords. This nondeterminism opens door for attacks as an adversary is only required to craft inputs that fall into the subspace of authentic inputs and does not need to exactly or almost match a genuine input. Signals crafted with predictive models did not exactly match the time-domain values of the authentic signals, but did successfully bypass the current state of the art in matching mechanism (Sadeghi *et al.* (2017) and section 4.2). Furthermore, even if an adversary is not aware of the mental task used for authentication (e.g. song recitation, imagined movement, closed eye resting) and provides her own brain signal, they can still get recognized as a valid user (Johnson *et al.*, 2014).

The second factor behind the intrinsic claim comes from the point that brain signals can only be collected at close proximity of user from their scalps and furthermore she would be aware of the procedure as it requires wearing a sensor on the head. In other words, brain signals cannot be captured remotely and without user's explicit knowledge (unlike face, iris, voice, and fingerprint). In addition, human activities do not leave a brain signal trace as it does in case of fingerprint, hair and skin. While these points are completely true regarding collectability of brain signals based on the current and foreseeable technology, however they do not imply intrinsic liveness property. In the simplest example, if a user records her own brain signals and later on feeds it to the biometric system, it would get authenticated without any problem.

But in reality the brain signal should not pass the liveness test as the *timely* property has been violated. In another scenario, we observe that recorded signals from one subject can get recognized as another subject while of course the timely property does not hold (for example check Sohankar *et al.*, 2015a; Johnson *et al.*, 2014). In more complex scenarios, as demonstrated in (Sadeghi *et al.*, 2017) and section 4.2, predictive models are capable of generating signals which can closely mimic behavior of genuine brain signals (bypass liveness guarantee) and furthermore impersonate users (bypass authentication guarantee).

2.1.4 Brain Liveness Problem

Based on close examination of literature and discussions in the preceding section (2.1.3), there is no strong basis to back the assumption of intrinsic liveness property for brain signals. Furthermore, the few challenging works Maiorana *et al.* (2013); Sadeghi *et al.* (2017); Gui *et al.* (2016) have provided evidence that brain signals similar to all other traits are vulnerable to presentation attacks and would require liveness detection methods for ensuring security and robustness. As the problem of brain liveness has been largely neglected in literature, there exists a gap with virtually no body of work on presentation attacks and liveness detection for brain signals. The mentioned works Sadeghi *et al.* (2017); Gui *et al.* (2016) focus on replay detection (through similarity checking with signals in system) as a solution for brain liveness which can be effective against manipulated inputs (i.e. noise addition attacks). However, as my work showed(Sadeghi *et al.* (2017) and section 4.2) such approaches are not as effective against synthesized inputs which are categorically different from set of signals in system database. In this thesis, I attempt to take the first steps toward bridging the gap by studying presentation attacks and liveness detection for brain signals due to the following motivating reasons.

1. There exist extremely limited work on brain liveness because of the inaccurate presumed intrinsic liveness which needs to be addressed.

2. Brain liveness is a more challenging problem compared to other human inputs since brain signals are not human perceivable, and lack evident and indisputable biological features which would prove their liveness (more details in section 2.1.7).

3. Brain signals possess features which makes them an appropriate candidate for ensuring liveness which in turn incentivizes us to design methods for enhancing its robustness against attacks. Universality, chaotic nature, and extremely fast response time are few of these features (more details in section 2.1.6.

I studied brain liveness detection from two different approaches. First, detecting presence of machine artifacts which are byproduct of the attack generation process. Second, Learning models using computational EEG features which can distinguish between genuine and fake signals. Based on these approaches, I proposed two set of solutions: *model-aware* and *model-agnostic*. In model-aware solutions, already known attack creation models/methods (e.g. noise addition, predictive, generative) are analyzed to determine artifacts that are associated with that model and then detect them in the input signals (details section 2.5.1). In model-agnostic solutions, attack creation process is not of emphasize, and instead learning models are trained using computational (e.g. frequency, wavelet, power, auto-regressive) features to classify input signals into EEG or Attack classes (details 2.5.2). I comprehensively evaluated these two approaches against different attack types to better understand their trade-offs (details section 2.8). I achieved error rates of less than 1% even in face of new (unseen) attacks which shows the generalizability of these approaches.

2.1.5 Preliminaries

Definitions

It was not until 2016, when International Organization of Standardization (ISO) and the International Electro-technical Commission (IEC) in a manuscript titled *Biometric presentation attack detection* (ISO, 2016) provided formal definitions, which we will use as base in this work, and customize them for our work.

- **Liveness:** "quality or state of being alive, made evident by anatomical characteristics, involuntary reactions or physiological functions, or voluntary reactions or subjects behavior." (ISO, 2016).

- **Liveness Detection:** "measurement and analysis of anatomical characteristics or involuntary or voluntary reactions, in order to determine if a biometric sample is being captured from a living subject present at the point of capture." (ISO, 2016).

- **Presentation Attack:** "presentation to the biometric data capture subsystem with the goal of interfering with the operation of the biometric system." (ISO, 2016). Examples of adversarial inputs mimicking valid human traits are replaying/generating voice commands or brain/heart signals, gummy finger with valid user fingerprints on it, displaying user pictures (printed or on screen) to a face recognition system, etc.

Liveness detection goal is to verify two properties: *live property:* input has been sensed from a live human subject, *timely property:* input has being sensed at the current point in time and is not a replay of an previous recording.

In this thesis, we focus on the first property of Liveness Detection (live input), distinguishing between brain signals captured from live human subjects and artificially generated signals. Any signal captured from a human scalp with brain activity under normal situations (e.g. awake, asleep, etc.) or abnormal situations (e.g. coercion, coma, etc.) is considered to be a *live* input. Signals that are not captured from a human scalp and have been completely/partially generated or manipulated in an artificial manner is considered to be *adversarial* input and not *live* input.

With respect to second property (timely input), determining if brain signal belongs to current point in time or past, the challenge lies in ground truth since as of now there is no standalone method to determine age of a brain signal. This issue exists also in other biometric traits; given a audio voice, heart signal, or even a face image without context one can not state when the trait was captured. For brain signals context be added in two way. First, by presenting random number of stimuli at random times to subject, and then attempting to detect its impact on brain signal. Second, capturing one or more secondary traits which do have correlation with brain signal.

There is a body of work on event detection for brain signals for biometric and medical application, which in nature is similar to detecting *timely* property based on context for brain liveness. Therefore, this work focuses on the detecting *live* property of brain signals using software solutions to bridge the gap in literature. Furthermore, in one of my works (Sadeghi *et al.*, 2016a), human subjects were presented with stress stimuli (horror movie scenes) and the increase in nervousness level was detected in brain and heart signals. Such context-based approaches have shortcomings such as need for presenting stimuli, extra hardware, user involvement, usability reduction, and non-universal impact on users which software-based approaches are free of them. However, software-based approaches are more challenging as no new information in form context or another sensed trait is not available, and decision making should be done solely based on the received brain signal.

Presentation attacks has also been refereed to as *spoofing attacks* in literature, and similarly liveness detection has been termed as *presentation attack detection*. I will be using these terms interchangeably in the rest of the paper, and the same goes with *human input* and *trait*.

Problem Statement and System Model

The brain LD problem can be formulated in the following manner: In a systems which receives brain signal as an input, how can we determine if the signal has been captured from a live human subject at current point in time? In other words, how to verify *live* and *timely* properties of input? It is noteworthy that detection of timely properties can only be done through utilizing context (e.g. providing random stimuli to subject and detecting its expected response) and as of know there is no standalone approach to detect timely property for human traits.

Figure 2.1, shows the system model for the problem in case of brain-based biometric systems, where an adversary provides adversarial input via brain sensor which goes through matching process for authentication and LD. The input will first be compared with stored traits for authentication purpose. If the input did match the claimed user's traits, it will be then checked for liveness, otherwise will be rejected. In the final stage, access to system will be granted only if input liveness is verified and otherwise access will be denied.

Liveness Detection Methods

There exists an extensive body of work an designing novel attacks and LD methods for traits such as fingerprint, face, iris, and voice (check anti-spoofing handbook (Marcel *et al.*, 2019), or surveys on LD for face recognition (Ramachandra and Busch, 2017), fingerprint (Marasco and Ross, 2015), iris (Czajka and Bowyer, 2018), and voice recognition (Wu *et al.*, 2015a)). The proposed LD methods in literature can be generally categorized into three approaches.

1. **Software-based**: extracting features from the input trait which can be evidence of its liveness (e.g. breath noise in voice)
2. **Context-based**: presenting a random challenge/stimuli to user and detecting its expected response, similar to role of CAPTCHA in detecting online bots (e.g. asking user to blink, twist finger, change voice tone)
3. **Hardware-based**: capturing another trait from user (e.g. body heat map, skin electric resistance, heart/brain signal)

With respect to robustness, context-based and hardware-based methods are more secure but less usable due to increased user involvement and need for extra sensor. In context-based methods, by randomizing type, time and number of stimuli/challenges presented, makes it impractical to craft forged traits in offline manner. Moreover, adding the expected response in real-time would be extremely challenging although not impossible. In hardware-based methods, the primary idea is to create more challenges for adversary as it now requires to craft either more articulated inputs (e.g. gummy finger with electric resistance similar to human body) or several inputs (e.g. face and brain).

Software-based methods do not reduce usability but at the same time do not receive extra help/information in form of context (for current trait) or an additional trait, which renders software-based methods to be more challenging then the other two. In this work on brain liveness, we specifically focused on software-based approaches since there exists a body of work on event/stimuli detection in brain signals (including one of mine (Sadeghi *et al.*, 2016a)) for biometric and medical applications which in nature is similar to context-based LD methods (Exarchos *et al.*, 2006; Brigham and Kumar, 2010; Chuang *et al.*, 2013; Abo-Zahhad *et al.*, 2016; Vidyaratne and Iftekharuddin, 2017; Zeng *et al.*, 2019; Nakanishi and Maruoka, 2019; O'Shea *et al.*, 2020). However, since there is a gap in literature regarding liveness detection methods which do not rely on context, and could detect presentation attacks using the input itself, I targeted software-based approached for brain LD.

Authentication & Liveness Detection Scenarios

To better highlight why both authentication and liveness detection are required, we will compare two scenarios. In the first scenario, when an adversary crafts a gummy finger to impersonate a legitimate user, on one hand authentication will fail in detecting the attack since the fingerprint reading is exactly or extremely close to the user's fingerprint signature (provided during registration). On the other hand, Livneness Detection (LD) would be successful since it will detect that the input is not coming from a live human finger and instead a gummy finger, so it violates the *live* property of liveness. In another scenario where an human adversary provides her own fingerprint to impersonate a user, LD will not be useful since the input indeed posses both the live and timely properties. However, authentication will succeed in preventing this attack since the input will not match the user's signature and will be rejected. As mentioned before, each of the two security guarantees alone can not prevent all types of attacks and they are need in conjunction for a robust system.

Brain Signal (EEG)

In this work, we will studying liveness for Electroencephalography (EEG) signals captured from human scalp surface area. These signals record the voltage difference between the scalp and some base (usually ear) and are in the range of 50-100 micro (μ) volts. EEG signals are generally decomposed into five frequency bands (delta, theta, alpha, beta and gamma) which correspond to different brain activities.

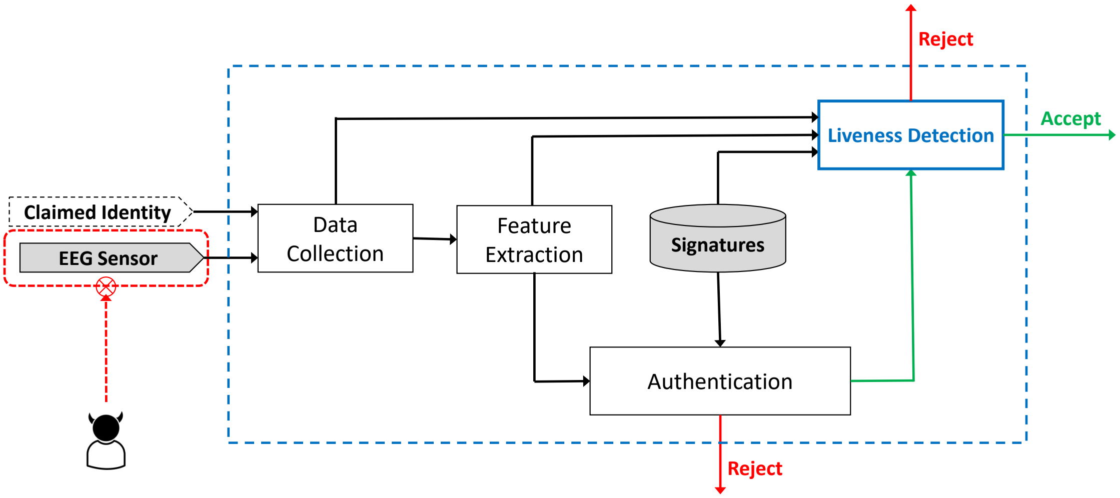


Figure 2.1: System Model for a Brain-based Authentication System with Liveness Detection

2.1.6 Why Brain?

Brain signals are equipped with a set of features, making it an ideal option for LD. While some of these features might also be available in other human inputs, brain is unique since it has them accompanied together. I will focus on features which are related to LD task, and not authentication task (such as uniqueness, for those features check (Sohankar *et al.*, 2015a)).

Universal: Human beings possess their brains until they are alive and with death of the brain, body can not live on without artificial care (bra, 2021). In several jurisdictions, brain death is considered as a sign of legal death (Jones *et al.*, 2018). While one might lose their finger, arm/leg, voice, or eye but their brains would be functioning and can still interact with systems utilizing brain signals. Therefore, brain signal allows for more inclusive systems which are also usable for individuals with disabilities.

Hands Free: In scenarios, where one is busy with other tasks and might not be able to use and move their face, hands, and voices (such as driving), brain signals would allow for pervasive systems without adding to user tasks and reducing usability. Furthermore, this characteristic can be used for continues monitoring instead of one-time initial checking.

Collectability: While an adversary can easily collect many human inputs (face image/video, voice, etc.) from online resources, capture from distance (taking image/video, recording voice), or extract from collected media (e.g. extracting iris and fingerprint from images), in case of brain adversary needs to invade user's personal space. Brain signals are recorded using wearable sensors and they can be used without user's knowledge or consent.

Chaotic Nature: The chaotic nature of brain signals renders them more challenging to be replicated or mimicked compared to other human inputs which have clear features and patterns. While there are some works on generative/predictive models for brain signals (Samanta, 2011; Sadeghi *et al.*, 2017), they are computationally more expensive due to brain's chaotic behavior.

Involuntary Reaction: Beside the voluntary actions, brain signals captures involuntary reactions of the user which can be utilized in context-based approaches for LD, where a stimuli is presented to user and its expected impact on brain signals are checked.

Response Time: Brain is the first part of the body which reacts to external stimuli in generally less than 300 ms (Mulholland *et al.*, 1976), and afterwards the impact can be seen in other organs. For example, stress can be detected using brain signals as fast as 250 ms, while it takes 3-4 s until it impacts heart signal (Sadeghi *et al.*, 2016a). This fast response time makes it more challenging for adversaries to add the expected impact of a stimuli (presented during context based LD methods) to their adversarial input in such a short window.

2.1.7 Brain Liveness Challenges

Liveness Detection for human inputs is not a trivial task as it is an arms race between adversary and the system where the attack space is extremely large to be fully explored and understood. On top of that, LD for brain signals face extra challenges. **Perceivability:** Contrary to other human inputs, brain signals are not human perceivable and do not carry any semantic properties. Furthermore, brain signals morphology is similar to random noise and even for an human expert distinguishing between genuine and fake brain signals would not be an easier task compared to machines. While a gummy finger, face masks, and printed images of face/iris are easily noticeable to an human observer, fake brain signal can not be detected on the spot.

Even in case of DeepFake (net, 2018d; Güera and Delp, 2018) (seemingly real but fake videos), the adversarial video should follow a wide range of semantic properties (lips move during talking, face movements should be consistent, regular blinking, normal body pose, etc). Lack of these expected semantic properties will lead to suspicion and detection of DeepFake videos. However, brain signals do not carry such semantics which makes detecting fake signals more challenging.

Biological Features: To best of our knowledge, brain signals do not possess any significant biological features such as PQRST complex in heart signals. Features such as P300 in braing signals are dependent on context and not characteristic of brain signals. This lack of discriminating features means brain LD would become more complicated and there would not be an definite/deterministic measure for determining authenticity of brain signals. This inherent non-determinism introduces a fundamental uncertainty in decision making for brain LD, which given an incoming brain input one can not be completely confident in accepting it as authentic brain signal.

To overcome this challenge, scholars need to turn into generic computational features (e.g. Fourier transform, Wavelet transform, Autoregressive Coefficient, etc.) which have shown in experiments that can be useful in distinguishing real and fake signals. A shortcoming of this approach is it's dependency on dataset and the problem of generalizing, which means features which are useful in one dataset, might not lead to high performance in another dataset. Another approach would be detecting machine artifacts which are created during the process of synthesis of fake signals. This approach is model-dependent as by changing the model for fake signal generation, the artifacts can change. Combining the two approach, and utilizing both computational features and machine artifacts could allow for more successful brain LD.

2.1.8 Emerging Challenges

With COVID global pandemic in 2020, majority of social interactions migrated from physical world to digital and remote realm, which might not fully reverse even in case of overcoming the pandemic. University exams, business interviews, and governmental meetings are now done remotely which opens door for impersonating attacks (especially considering success of deepfake videos (net, 2018d; Güera and Delp, 2018)). Therefore to protect the integrity of our remote communications which social functions are currently dependent on it, and also considering the astronomical increase in the volume of remote communication, usable and scalable liveness detection methods should be designed.

Emergence of deepfake videos (net, 2018d; Güera and Delp, 2018) which can successfully mimic human face and voice in form of video poses threats for face and voice recognition systems. Moreover, it gives hint that with advances in Artificial Intelligence (AI), it should be assumed that all content/data can possibly be forged artificially, and there is a need for liveness checking in all domains (e.g. text, audio, video, signal, art, etc.) to distinguish human generated works from those of machine. Already there exists works which perform well in generating different types of text (Brown *et al.*, 2020), music (Dhariwal *et al.*, 2020) and art (Foster, 2019).

In a recent work, Wang *et al.* (2020b) have proposed a generative model for synthesis of a human talking head which could significantly reduce the data communication required for video calls. Such an approach could also possibly allow for creation of real-time deep fake videos which could impersonate an individual in interactive scenarios such as remote meetings, interview and exam. With ongoing advances in vision and natural language processing fields, we will be observing fake avatars which fool human observers, and might not also be easily detectable using vision approaches. A potential solution is using brain for LD and detecting the correlation between user's behavior and face movements with their brain signals.

2.1.9 Summary of Contributions

In summary, in this chapter I make the following contributions:

- First work to comprehensively study liveness detection problem for brain signals, and propose two novel solution approaches (model-aware and model-agnostic) based on detecting machine artifacts in adversarial inputs and finding computational features in authentic inputs.

- Evaluate the solution approaches against different types of presentation attacks (manipulated input (noise addition) and artificial input (predictive and generative models)) and achieving error rates of less than 0.01%.

- Constructing the first publicly available attack databases for brain signals' liveness detection to allow for bench-marking and comparison of future works (to be available on our lab website). While there are public presentation attack databases for other human traits (i.e. fingerprint, face, finger vein, palm vein, and voice) (Chingovska *et al.*, 2019), there had been a lack in case of brain signals.

2.2 Related Works

In table 2.1, I have compared performance of 27 liveness detection methods for face, fingerprint, iris and voice with each other and the current work. In the second column of table, I have mentioned the database or competition which was used to evaluated these methods, as performance is significantly dependent on the dataset used. A challenge in LD is generalizability of methods, and a method which performs well on one type of attack might perform poorly against other ones.

From table 2.1, it can be seen that the subject pool used in biometric liveness is generally small (best case 504), and there is need for large datasets which would allow for bench-marking and more accurate comparison. The Half Total Error Rate (HTER) range between 0-29.78% with average of 7.32% which still needs more research and improvement. The four traits beside brain, have had public datasets and few series of competition (e.g LivDet, LivDet-Iris, ASVspoof) which aids and encourages further research. However brain liveness lacks such tools and as the first step this thesis will publish it's dataset on our lab website (impact.lab.asu.edu) in the hope of other works improving brain liveness detection methods.

The studied presentation attacks can be classified in two main groups; synthetic and manipulative (check third column in table 2.1). In synthetic attacks, adversary artificially crafts an input such as gummy finger, prosthetic eye and iris contact lens, face mask and synthetic voice/brain signals. In manipulative attacks, a genuine user trait is changed and presented as input such as displaying a printed image of face or noise addition to user's voice/brain signal. A replay attack, where an unchanged user

Trait	Attack Type	Attack Approach	Subjects	Database/Competition	Method/Team	FAR	FRR	HTER
Face	Print & Display (photo & Video)	Manipulation	50	REPLAY-ATTACK (Chingovska *et al.*, 2012)	(Chingovska *et al.*, 2012)	NA	NA	17.17
Face	Print & Display (Video)	Manipulation	50	CASIA (Zhang *et al.*, 2012)	(Raghavendra and Busch, 2014)	NA	NA	10.21
Face	Print & Display (Video)	Manipulation	50	CASIA (Zhang *et al.*, 2012)	(Benlamoudi *et al.*, 2015)	11.39	11.39	11.39
Face	Print	Manipulation	50	PRINT ATTACK (Anjos and Marcel, 2011) — Competition on Counter Measures to 2-D Facial Spoofing Attacks - 6 Teams - 2011 (Chakka *et al.*, 2011)	Three Teams (CASIA, IDIAP, UOULU) (Chakka *et al.*, 2011)	0.00	0.00	0.00
Face	Print & Display (photo & Video)	Manipulation	50	REPLAY-ATTACK (Chingovska *et al.*, 2012) — 2nd Competition on Counter Measures to 2-D Facial Spoofing Attacks - 8 teams - 2013 (Chingovska *et al.*, 2013)	Two Teams (CASIA, LNMIIT) (Chingovska *et al.*, 2013)	0.00	0.00	0.00
Face	Print & Display (Video)	Manipulation	55	OULU-NPU (Boulkenafet *et al.*, 2017b) — Competition on Generalized Software-based Face Presentation Attack Detection in Mobile Scenarios - 13 teams - 2017 (Boulkenafet *et al.*, 2017a)	GRADIANT(extra) over 4 evaluation protocol (Boulkenafet *et al.*, 2017a)	2.6 - 7.1	1.9 - 5.8	2.5 - 10.5
Finger Print	Play Doh Finger & Cadaver Images	Synthetic & Manipulation	33	WVU04 (Abhyankar and Schuckers, 2004)	(Abhyankar and Schuckers, 2004)	0.00	0.00	0.00
Finger Print	Images of Gelatin/Clay	Manipulation	23	Hong Kong (Moon *et al.*, 2005)	(Moon *et al.*, 2005)	0.00	0.00	0.00
Finger Print	Play Doh Finger & Cadaver Images	Synthetic & Manipulation	33	WVU01 (Derakhshani *et al.*, 2003)	(Derakhshani *et al.*, 2003)	11.11	11.11	11.11
Finger Print	Silicon, Gelatin, Play-Doh Images	Synthetic	464	LivDet 2009 (Marcialis *et al.*, 2009) - 4 Teams	Dermalog (Marcialis *et al.*, 2009)	5.40	20.10	12.75
Finger Print	Silicon, Gelatin, Play-Doh, Latex, Wood Glue Images	Synthetic	256	LivDet 2011 (Yambay *et al.*, 2012) - 5 Teams	Federico (Yambay *et al.*, 2012)	24.50	26.60	25.55
Finger Print	Body Double, latex, Play-Doh, wood glue, gelatine, ecoflex, modasil Images	Synthetic	414	LivDet 2013 (Ghiani *et al.*, 2013) - 10 Teams	UniNap1 (Ghiani *et al.*, 2013)	14.62	11.96	13.29
Finger Print	Ecoflex, gelatin, latex, wood glue, liquid Ecoflex, RTV, Play-Doh, Body Double, OOMOO images	Synthetic	284	LivDet 2015 (Mura *et al.*, 2015) - 9 Teams	nogueira (Mura *et al.*, 2015)	4.26	5.26	4.76
Finger Print	Ecoflex, gelatin, latex, wood glue, liquid Ecoflex, Body Double images	Synthetic	n.a.	LivDet 2017 (Mura *et al.*, 2018) - 12 Teams	JLW B (Mura *et al.*, 2018)	5.05	4.40	4.73
Iris	Print	Manipulation	54 (27 subjects * 2 eyes)	ATVS (Ruiz-Albacete *et al.*, 2008; Galbally *et al.*, 2012b)	(Raghavendra and Busch, 2014)	NA	NA	0.00
Iris	Print & Contact Lens	Synthetic & Manipulation	>630 eyes	LivDet-Iris 2013 (Yambay *et al.*, 2014) - 3 Teams	Federico (Yambay *et al.*, 2014)	5.72	28.56	17.14
Iris	Print & Contact Lens	Synthetic & Manipulation	>900 eyes	LivDet-Iris 2015 Yambay *et al.* (2017b) - 4 Teams	Federico Yambay *et al.* (2017b)	5.48	1.68	3.58
Iris	Print & Contact Lens	Synthetic & Manipulation	>793 eyes	LivDet-Iris 2017 (Yambay *et al.*, 2017a) - 3 Teams	Anon1 (Yambay *et al.*, 2017a)	14.71	3.36	9.04
Iris	Print, Contact Lens, Display, Cadaver, Prosthetic	Synthetic & Manipulation	>>42subjects	LivDet-Iris 2020 (Das *et al.*, 2020) - 3 Teams	USACH/TOC	59.10	0.46	29.78
Iris	Print	Manipulation	200 eyes	MobILive 2014 (Sequeira *et al.*, 2014) - 6 Teams	IIT Indore (Sequeira *et al.*, 2014)	0.00	0.50	0.25
Voice	Synthetic Voice	Synthetic	283	Created by Authors	(De Leon *et al.*, 2012)	2.50	2.50	2.50
Voice	Converted Voice	Manipulation	504	subset of NIST SRE 2006 (NIS, 2006)	(Wu *et al.*, 2012)	9.29	9.29	9.29
Voice	Converted & synthetic Voice	Synthetic & Manipulation	106	ASVspoof 2015 (Wu *et al.*, 2015b) - 16 Teams	(Patel and Patil, 2015)	1.211	1.211	1.211
Voice	Replay Voice	Manipulation	42	ASVspoof 2017 (Kinnunen *et al.*, 2017) - 49 Teams	(Lavrentyeva *et al.*, 2017)	6.73	6.73	6.73
Voice	Converted & Synthetic	Synthetic & Manipulation	78	ASVspoof 2019 (Todisco *et al.*, 2019) (Logical Access scenario) - 63 Teams	T05 (Todisco *et al.*, 2019)	0.22	0.22	0.22
Voice	Replay Voice	Manipulation	78	ASVspoof 2019 (Todisco *et al.*, 2019) (Physical Access scenario) - 63 Teams	T28 (Todisco *et al.*, 2019)	0.39	0.39	0.39
Brain	Generated & Noise Added Signals	Synthetic & Manipulation	106	Genertated by authors	Current Work	0.00	0.00	0.00

Table 2.1: Comparison of Liveness Detection Methods for Different Human Traits. FAR Stands for False Accept Rate, FRR Is False Reject Rate, and HTER Is Half Total Error Rate Which Is the Arithmetic Mean of FAR and FRR.

trait is presented is also considered as manipulative attack.

2.3 General Problem

In this section, I formulate the liveness problem in an abstract manner and discuss it in relation to presentation attacks where an adversary presents malicious input to interfere with system's normal operation and how to verify if inputs belong to a specific source or not. This formulation holds not only for biometric systems, but also for systems which operate base on inputs from physical world such as cyber-physical systems.

A System receives an input set (I) which is assumed to belong to a specific source S. System's behavior is determined by a decision making function, F, which receives the input set I and possibly another set of internal parameters (P), and generates an output:

$$\begin{gathered} F(I,P) = Out, \\ I = \{i_1, i_2, .., i_n\}, \\ P = \{p_1, p_2, .., p_k\} \end{gathered} \tag{2.1}$$

Output of this function (Out) controls system's actions. Since F and P are fixed, an adversary can manipulate system's behavior by presenting an input which seems to belong to the source S. There is three assumptions here: 1) Function F is fixed, 2) Parameters P are fixed, 3) adversary does not have access to source S. The first two assumption suggests a threat model where an adversary can not change the system's internal settings and can only provide input, however she might have knowledge of the function F and parameters P. There can be threat models where adversary can also change F and P which suggests attacker has gained control over the system and has other means of changing system behavior beside simply providing malicious input to reach her goals. With regard to the third assumption, if the adversary has access to source S then she basically can provide any desired input and there is no challenge in changing system's behavior. In practice, not only the adversary usually does not have access to the source S, but in many cases the system and it's designers do not have such access either, and they only have access to some samples from that source. Generally, adversary has access to some number of samples from source S and constructs another source, $S^{'}$, which is supposed to mimic source S. Estimating distance and difference between sources S and $S^{'}$ is in itself another challenging task and out of the scope of this paper.

The mitigate adversary's attacks, another function, G, should verify if system input I belongs to source source S:

$$G(I, S) = \begin{cases} 1, & I \in S \\ 0, & otherwise \end{cases} \tag{2.2}$$

To be more precise, the exact specification and distribution of source,S is usually not known and function G should instead work on a set of samples from the source (A):

$$A = \{a_i | a_i \in S\}$$

$$G(I, A) = \begin{cases} 1, & I \in S \\ 0, & otherwise \end{cases} \tag{2.3}$$

In a more complicated scenario, inputs also have temporal dimension (I_x) and it should be verified that if it does match time t, which is generally the current point in time but it can also be a time in past.

$$G(I_x, A, t) = \begin{cases} 1, & I \in S \wedge |x - t| \leqslant \varepsilon \\ 0, & otherwise \end{cases} \tag{2.4}$$

The value of ε can be set based on system requirements or trade-off between robustness and performance.

2.3.1 challenges

To resolve the described problem, several challenges regarding the two functions, F & G, and source S should be considered.

First, In many domains (including cyber-physical), due to nature of the inputs and the expected output (e.g. autonomous vehicle & voice recognition), system's decision-making function, F, is non-deterministic. It means that type of the function and its parameters can not be determined analytically, and are decided by other approaches (e.g. experimental, probabilistic, heuristics, etc.). Therefore, there is an inherent uncertainty and error in mapping between input and output. Furthermore, the function F might not have a straightforward format (e.g. function representing a deep neural network) and therefore it cannot be properly analyzed. These functions are similar to cryptographic hash functions in aspects such as they cannot be reversed, are not differentiable, and small changes in input (even 1 bit) can lead to dramatic changes in output. These properties and uncertainties opens doors for adversaries to launch presentation attacks using crafted inputs to reach their desired output or at least make the system malfunction.

Second, function G which attempts to prevent presentation attacks is also non-deterministic itself as the true distribution of source S is unknown in most cases. Hence, no entity (human or machine) can determine with 100% confidence if the input indeed belongs to the source S or not. So while adding function G to system helps with preventing attacks and improves confidence in output of function F, it also means adding another non-deterministic function to system which is however unavoidable in most cases.

Third challenge is with source S as it can be unknown and the knowledge about it can be limited to some number of observed samples (A, Eq 2.3). Estimating the true distribution of S based on the set of observations (A), is not a trivial task in itself and there can more complex scenarios. As mentioned above, there can be a case where inputs' temporal dimension should also be verified (Eq 2.4). In a more complicated case, Source S is consisted of a set of sources itself, and each of these sources have their own distinct distribution.

$$S = \{s_1, s_2, .., s_m\} \tag{2.5}$$

In other words, there is no concrete distribution for set S itself and it depends on the distributions for its members. If these individual distributions are similar to each other then there is more chance of estimating some sort of general distribution for set S. However, if they are not similar or there is zero or limited knowledge about these individual distributions, estimating distribution for set S becomes even more challenging. So the knowledge about S would be limited to some number of observations (samples) for some of the distributions and then the problem of answering if input I belongs to set S would become extremely challenging.

$$\begin{aligned} & A = \{a_i | a_i \in s_j\} \\ & G(I_x, A, t) = \begin{cases} 1, & I \in S \land |x - t| \leqslant \varepsilon \\ 0, & otherwise \end{cases} \end{aligned} \tag{2.6}$$

An example of such scenario is brain signal liveness; there is limited EEG recordings (a_i) available from limited number of individuals' brain (s_j) compared to the set of world population brains (S). Furthermore, there is limited knowledge on brain dynamics and distribution for each individual and we are faced with the problem of determining if a given signal belongs to class of human brain signals and to the current point in time.

2.4 Adversarial Cycle

To approach the general problem discussed in section 2.3, I suggest an workflow which is usable for both attack and defense purposes for liveness detection of inputs from physical world as seen in Figure 2.2.

The workflow operates as follows: First, one input from the set of system inputs are chosen. Second, generate fake inputs which mimic the chosen input. There is generally two approaches for faking data; artificial synthesis or manipulating a genuine input. Third, check the quality of the fake input and its resemblance to genuine inputs. Based on the domain there can be different methods but two general approaches would be information theoretic (e.g. entropy analysis) or checking behavior in time and feature domain. Fourth, use the fake data to attack the system. If attack failed then either quality of the generated fake data should be improved (go back to step two) or change the input and choose another one to be faked (go back to step one).

If the attack was successful, then there is a need for a liveness detection method (the discussed function G in 2.3) for this input. In fifth step, a liveness detection method should be designed. In sixth step, attack the liveness detector designed in previous step. At this stage, two cycles could happen. If the attack fails, we need to go back to step two and create enhanced fake inputs and try again on attacking the liveness detector. This cycle is done by the adversary for increasing its chance of bypassing the liveness detector. If attacking the liveness detector had succeeded, then the liveness detector needs to be improved and then again be tested. This cycle is undergone by the defense side to find new ways of detecting attacks. These two cycles can continue for a specific number of time or as some threshold is met.

I will be using this workflow to study the brain liveness. In this chapter, the focus is on the second cycle (improving liveness detector) and two solution approaches are proposed and for each approach several different method has been studied. In section 4.2, focus is on the first cycle, improving fake data quality and attacking both the system and its liveness detector. Six different attacks with increasing complexity are designed and tested against 30 brain authentication configurations equipped with liveness detector in form of similarity checking in time and feature domain.

2.5 Solution Approaches

To tackle the brain LD problem, I propose two different approaches: *Model-Aware* and *Model-agnostic*. Building upon the discussions in challenges section 2.1.7, model-aware approaches focus on detecting machine artifacts of fake signals while model-agnostic approaches utilizes computational features to train models which can distinguish fake and real brain signals. The two approaches also can be combined into an hybrid approach where input is both checked for machine artifacts and also undergo classification based on computational features and afterwards the two outcomes are fused into final decision.

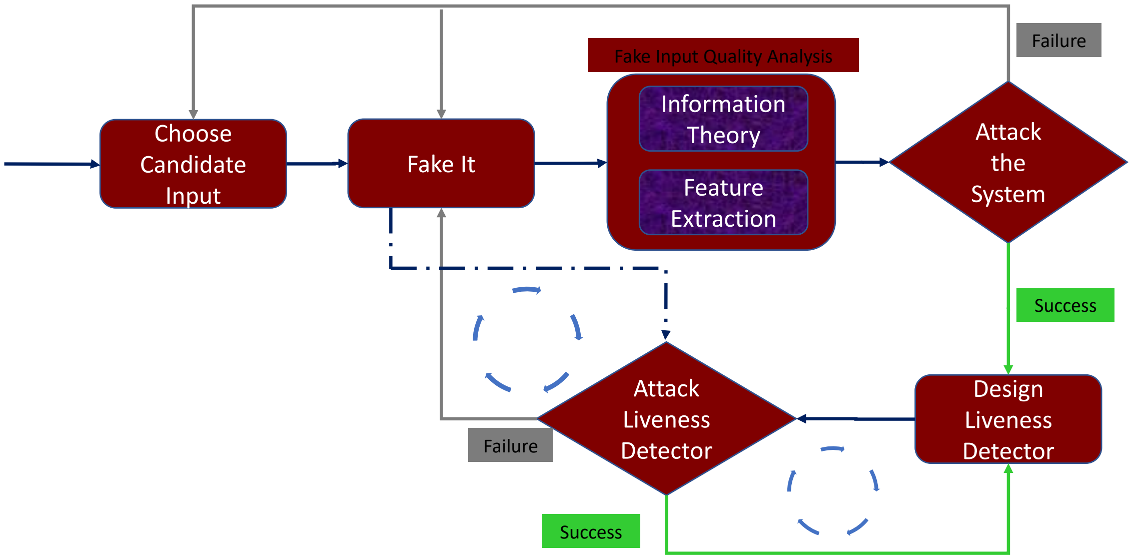


Figure 2.2: Adversarial Cycle for Liveness Detection in Cyber-physical Systems

2.5.1 Model-Aware Approach

Machine generated signals may have artifacts that are caused by certain fundamental properties of the machine. For example, in Generative Adversarial Nets, GAN (Goodfellow *et al.*, 2014) which are used to generate fake image signals, de-convolution operation performed at the generator can have checkerboard effect (Zhang *et al.*, 2019; Wang *et al.*, 2020a; Damer *et al.*, 2019). This effect introduces peaks at unwanted frequencies in the image spectrum. This is a fundamental property of the de-convolution operation as it is caused by the de-convolution window. It is extremely hard to completely get rid of checkerboard effect. The only possible way to reducing it to some extent is to use upsampling or overlapping windows (Shi *et al.*, 2016). Both those strategies lead to increased training time and poorer performance of the machine. In *model-aware approach*, I assumed that we have knowledge about the machine that is used to generate fake data. Although this is not an hard requirement and attacks generated with unknown models can have similar artifact types as the ones observed in the studied models. From the available knowledge, a pattern can be derived for the artifacts of a machine. The artifact pattern is then used to distinguish between live signal and machine generated signal. The main component of this approach is an artifact extraction algorithm that is machine specific. This is a drawback of this technique since often an accurate artifact extraction algorithm may not exist. I discuss the model aware approach for the learning models of Adaptive Neuro-fuzzy Inference System, ANFIS (Jang, 1993) and Generative Adversarial Nets, GAN (Goodfellow *et al.*, 2014)) (synthetic attacks) and noise addition (manipulative attacks). In both the cases, we can identify frequency domain artifacts. GANs are specifically designed for images. In our work we have adopted GANs to generate fake time domain signals. However, we observe similar checkerboard effect in the frequency domain features.

Machine Artifacts

Figures 2.3, 2.4 visualizes the artifacts observed in Discrete Wavelet Transform (DWT) feature vectors of attacks compared to brain signal. EEG feature vector is on the left side of the first row, the next two plots show Adaptive Neuro-fuzzy Inference System, ANFIS (Jang, 1993) attacks in time and frequency domain, and the last plot in first row is Generative Adversarial Nets, GAN (Goodfellow *et al.*, 2014) attacks. The second row shows feature vectors for white noise addition in frequency domain with Signal-to-Noise-Ratio (SNR) of -10,-1, 1 and 10. Finally, The third row shows the case for white noise addition in time domain with the mentioned SNR values.

In figure 2.3 feature vector is details coefficients in level one extracted using DWT with Symlet (*sym10*) wavelet. For GAN attack and the five time-domain attacks (ANFIS-time and four time noise), a significant amount of artifacts is introduced. Furthermore, the range of values for EEG features (row 1, col 1) is [-1,1], but in the time-domain attacks is increased to at [-5,5] to [-50,50]. For the five frequency domain attacks (ANFIS-frequency and four frequency noise), artifacts are in the form of several peaks with values in range of [-6,6] which go outside the expected range of EEG features [-1,1].

In figure 2.4 feature vector is approximation coefficients in level four extracted using Symlet (*sym10*) wavelet. In this case, significant artifacts which impact the shape and range of the features are observable. Only the ANFIS-time (row 1,col 2) is more similar to EEG features, however the peaks in the beginning and end are reversed.

Figures 2.5, 2.6 visualizes the artifacts observed in Power Spectral Density (PSD) feature vectors of attacks compared to brain signal for the two cases on window of size 500 and 160. Similar to the case with DWT details features (Figure 2.3), GAN and time-domain attacks have more artifacts compared to feature domain attacks. For ANFIS-time attack (row 1, col 3), a pattern has repeated for three times, and for lower noise levels (SNR 1 and 10) in time domain (row 3, col 3 and 4) the shape of features to some extent resembles EEG features but with with large amount of artifacts. Frequency-domain attacks have relatively low amount of artifacts mostly in form increased variation.

These figures show machine artifacts do impact feature vectors extracted from input signal and can be utilized to distinguish between EEG signal and attacks.

Signal Entropy

Table 2.2 compares the information content or entropy of real brain signals with fake signals crafted with generative model approach (GAN), and predictive model approach (ANFIS in time and frequency domain). We can see for approximate entropy only the ANFIS (time domain) signals can preserve the entropy, and the other two

	Approximate Entropy	Shannon Enteroy	Log Energy Entropy
Brain Signal	0.4565 (0.0448)	-207,130 (469,210)	505 (235.55)
GAN Signal	0.7917 (0.0521)	-4,237,700 (2,591,500)	1064.2 (95.59)
ANFIS (Time) Signals	0.4630 (0.0429)	-143,210 (325,260)	458.39 (245.92)
ANFIS (Frequency) Signals	0.7958 (0.0913)	-215,330 (832,930)	370.80(328.44)

Table 2.2: Entropy Comparison Between Brain Signals and Adversarial Inputs Created Using Generative Adversarial Nets (GAN) (Goodfellow *et al.*, 2014), Adaptive Neuro-fuzzy Inference System (ANFIS) Jang (1993) in Time-domain, and Frequency-domain. The Mean and Standard Deviation Is Shown in Each Cell (in Parentheses) and Were Calculated over 1000 Signals.

fake signals have much higher entropy. This is a sign of machine artifacts which can point out presentation attacks. For the Shannon entropy and log energy entropy, we observe a similar pattern of significant difference between real and fake signals. Therefore, entropy can be used as one of the indicator for brain liveness detection.

Furthermore, from this table we can observe that fake signals change the information content of the input but the current processing mechanisms are not sensitive to it since such fake signals do got accepted by brain-based authentication systems. This brings attention to revising how inputs from physical world are processed so that such fake data would have less chance in bypassing the system.

2.5.2 Model-Agnostic Approach

In most cases, knowledge about how the adversarial inputs have been crafted is not in hand as there are numerous methods to create them and new ones are developed frequently. Considering the extensively large search space, it will not be possible to cover all models which can produce inputs in that space, and their potential artifacts. Hence, we also studied *Model-Agnostic* approaches, which takes a more generalized approach in learning features which can be successful in distinguishing real brain

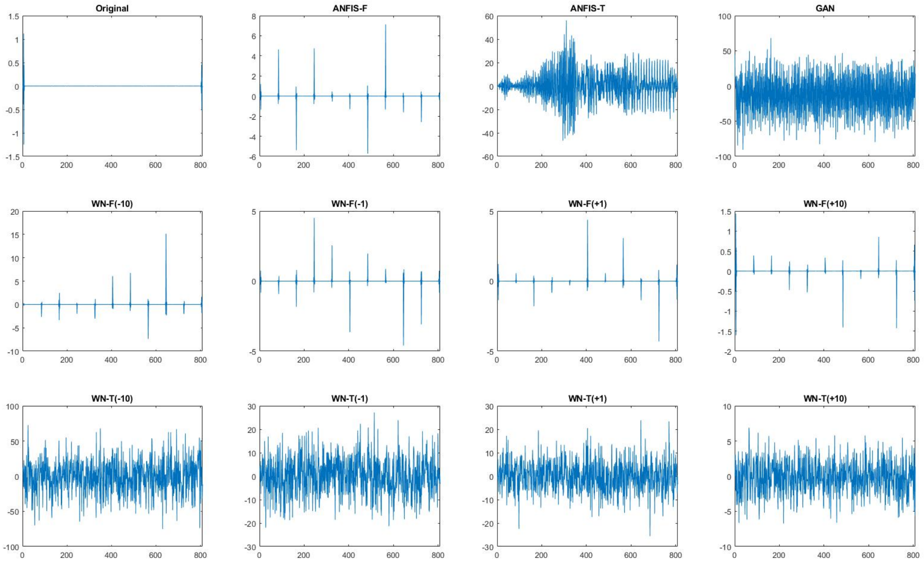


Figure 2.3: Machine Artifacts Visible in Details Coefficients in Level One of Symlet (*sym10*) Wavelet Features. WN-F and WN-T Stand for Signals with White Noise Added in Frequency Domain, and Time Domain Respectively. Numbers in Parenthesis Are the Signal to Noise Ratio (SNR).

signals from fake ones regardless of the model used in crafting them. Instead of directly focusing on machine artifacts (related to specific models) which can be a sign of adversarial inputs, in model-agnostic approach we rely on the appropriate combination of features and decision making functions that can label an incoming input. As discussed in section 2.1.7, to the best of our knowledge there is no definitive feature or characteristic for brain signals, and therefore, we turned into computational features. For the decision making functions, we focused on machine learning models which first go through training phase and then are expected to make decisions for future unseen cases. This approach is not limited to using machine learning models, and other types of decision making functions can also be utilized.

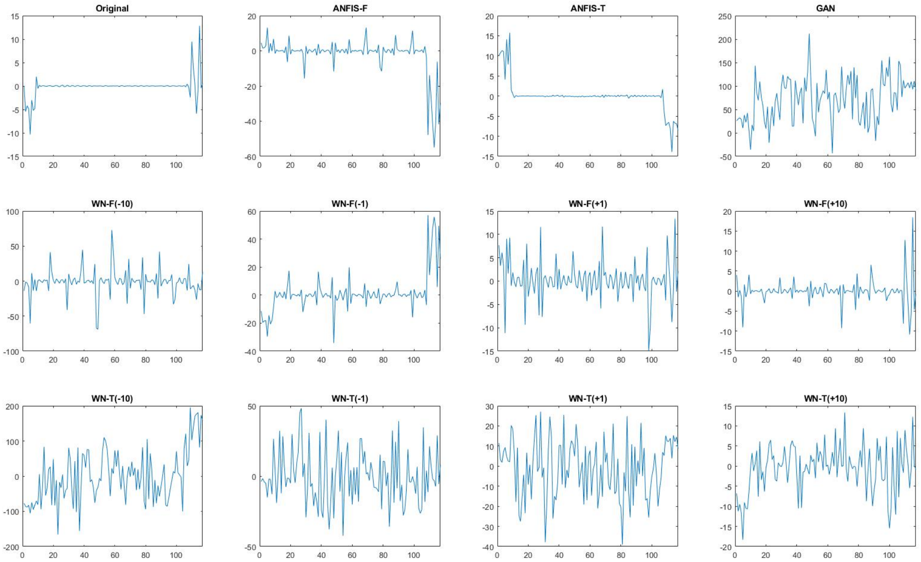


Figure 2.4: Machine Artifacts Visible Approximation Coefficients in Level Four of Symlet (*sym10*) Wavelet Features. WN-F and WN-T Stand for Signals with White Noise Added in Frequency Domain, and Time Domain Respectively. Numbers in Parenthesis are the Signal to Noise Ratio (SNR).

2.6 Threat Model

2.6.1 Attack Simulation

For each user, 42 different adversarial inputs (attack vectors) were crafted through predictive model or noise addition. For each user, out of the total six min available data (three signals of length two min), segments of 30s length from each of the three signals (total of 90s) are not used in training the model and is saved for testing (more details in section 2.7.2). Adversary has access to these test signals, and crafts attacks using them. Since the adversary knows the preprocessing step used in the system, it will first apply the 5th order Butterworth bandpass filter in range of 8-13 hz to them and then creates attack vectors. In summary, for each user 42 adversarial inputs of length 90s was created. The adversarial inputs were generated as below:

Noise Addition in Time Domain: Adversary creates 20 attack vectors by adding white Gaussian noise to the test data in time domain. Noise was added at 20 levels of Signal to Noise Ratio (SNR): $\{-10, -9, ..., -1, 1, ..., 10\}$.

Noise Addition in Frequency Domain: Adversary creates 20 attack vectors by adding white Gaussian noise to the test data in the frequency range of 8-13 Hz. Noise was added at 20 levels of Signal to Noise Ratio (SNR): $\{-10, -9, ..., -1, 1, ..., 10\}$.

Predictive Model in Time Domain: Adaptive Neuro Fuzzy Inference System (ANFIS) (Jang, 1993) was used for signal generation in time domain, which has been used in literature for generation of chaotic signals (Samanta, 2011). We did three-step ahead data point prediction based on ANFIS being trained on a window of 3s (480 data points). Then moved the training window three data point ahead, and re-trained and predicted three more data points. This process was separately done on each of the three 30s signal segments, to generate three new 30s signal segments, which are the attack vectors.**Generative Model in Frequency Domain:** ANFIS was used to generate signals in frequency domain in range of 8-13 Hz. Instead of having one ANFIS model, we had one for each of real and imaginary parts of the frequency values between 8-13 Hz (total of 12 models). Each model was initially trained on 30 sample points (i.e. frequency values of the 30s signal segment), did one step ahead prediction, then the training window was moved one data point ahead (which would include the newly predicted value), re-trained and again predict one new data point. Fast Fourier Transform (FFT) was used to got the frequency values, and after finishing the predictions, inverse FFT was applied to construct the time domain signal. This process was separately done on each of the three 30s signal segments, to generate three new 30s signal segments, which are the attack vectors.

2.7 Experiment Setup

I used a publicly available raw EEG dataset which contained three trails of two minute length from 106 subjects (Schalk *et al.*, 2004). EEG signals were sampled at 160 Hz using medical sensor with 64 channel electrodes while subjects opened and closed their right or left fist. Since commercial off-the-shelf sensors have less number of electrodes (generally 1-14), we decided to use signals from only one of the channels ("C3"), so that the proposed model would be usable for signals from any sensor and not depend on number of electrodes. As a pre-processing step, a 5th order Butterworth band pass filter in the range of 8-13 Hz was applied to raw signals.I comprehensively evaluated our model-agnostic approach under two protocols: *normal* and *unseen*

1. **Normal Protocol:** Training set includes samples from all categories of attack available in test set, but does not include all attack types within each category. For example, there will be samples from noise-addition attack category in training set, but not from all different noise levels (i.e. Signal-to-noise [SNR] ratio).

2. **Unseen Protocol:** Some categories of attack are not included in training set, and model gets tested against unseen attacks. For example, training set might only include noise-addition attack samples, but then tested against predictive model attacks.

2.7.1 Models and Feature Extractions

I benchmarked the attacks against 54 different machine learning configurations; nine feature extraction methods, and six models.I tested the models against different input lengths in range of 1 to 10 seconds. Feature extraction was either performed

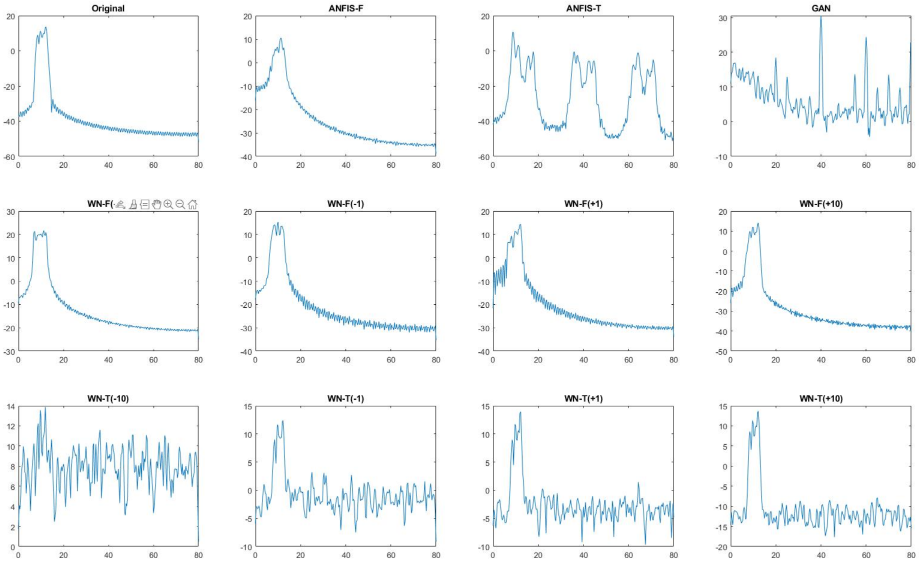


Figure 2.5: Machine Artifacts Visible in Power Spectral Density (Welch Method). WN-F and WN-T Stand for Signals with White Noise Added in Frequency Domain, and Time Domain Respectively. Numbers in Parenthesis Are the Signal to Noise Ratio (SNR).

once on the total input length or applied to one second segments of the input (160 data points), and then concatenated together to create feature vector. The following feature extraction methods were used:

1. **Fast Fourier Transform (FFT)**: Frequency values in range of 8-13 Hz.
2. **Statistical Analysis (SA)**: Mean, skewness, kurtosis, standard deviation, entropy, and range.
3. **Discrete Wavelet Transform (DWT)**: Fourth level approximation coefficients of Daubechies (*db1*) wavelet
4. **Auto Regressive (AR)**: Tenth order coefficients .
5. **Power Spectral Density (PSD)**: Calculated for the range of 8-13 Hz.

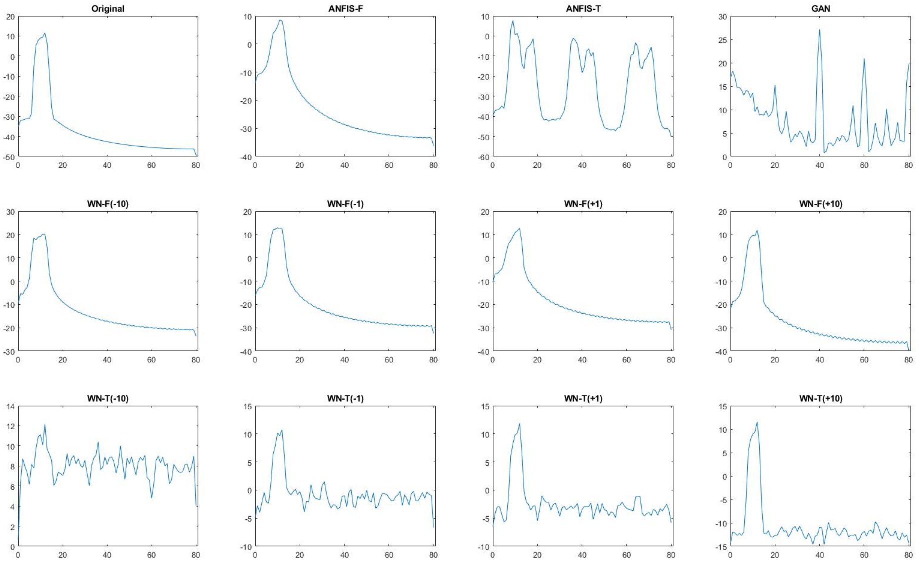


Figure 2.6: Machine Artifacts Visible in Power Spectral Density (Welch Method with 160 Points Window). WN-F and WN-T Stand for Signals with White Noise Added in Frequency Domain, and Time Domain Respectively. Numbers in Parenthesis Are the Signal to Noise Ratio (SNR).

6. **Combine (CMB):** All the above feature vectors were concatenated together.

7. **Principle Component Analysis (PCA):** Performed on the combine feature vector (CMB), and the 10 most significant components were chosen, which would explain 67% of the total variability. We also used 20 components (77% of variability) and 40 components (91% of variability) as features, but since the results were similar, we only reported the case with 10 components.

8. **DWT (sym10-L4)**: Fourth level approximation coefficients of Symlet *sym10* wavelet

9. **DWT (sym10-L1)**: First level details coefficients of Symlet *sym10* wavelet

I focused on lightweight models which are not computationally expensive and would not need large amount of training data. In this manner, mobile devices could locally

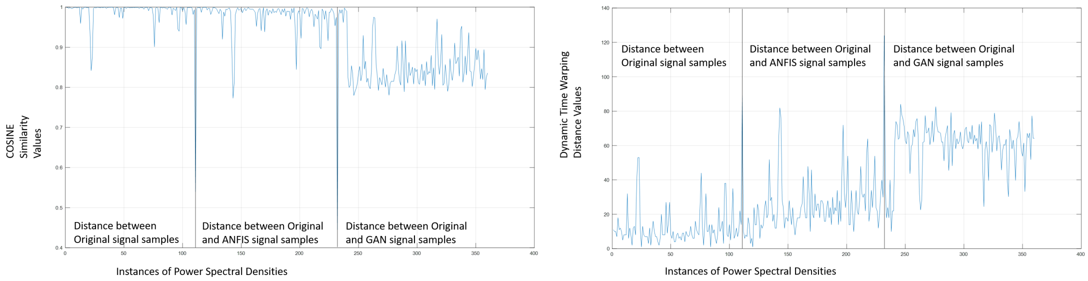


Figure 2.7: Comparison of Cosine Similarity and Dynamic Time Wrapping (DTW) Distance Between Power Spectra of Original, ANFIS Generated, and GAN Generated Signals.

train and test models without the need to share data or offload computation to cloud servers. The following six classifiers were used; Linear Discriminant Analysis (LDA), Naïve Bayes Classifier (NBC), Support Vector Machine (SVM) with RBF kernel, k-Nearest Neighbors (KNNs) and Ensemble with adaptive logistic regression (LogitBoost).

2.7.2 Training and Testing

Each trail signal (120 s) was divided into 12 samples of 10s length which considering 106 subjects and three trail per subject created total of 3816 samples. From each trail, the first nine samples were used in training (75% = 2862 samples) and the last three samples were saved for testing (25% = 954 samples). From each sample in the test set, 42 attack vectors were created (as described in section 2.6.1), so total of 40068 attack samples ($42 \times 954 = 40688$). Samples were labeled into two classes; *live* and *adversarial*.

For training, we had 2862 sample from the live class (75% of the EEG dataset), so for the adversarial class we randomly chose the same number of samples out of the generated attack samples. In other words, a random subset of size 2862 out of the 40068 attack samples were used in training (2.4%), and the rest were saved for

		Thresholds						
		mean	median	mean + std	mean + 2*std	mean + 3*std	99%	max
Distance	Euclidean	13.98%	20.54%	6.69%	9.33%	14.63%	13.53%	29.45%
	Standardized Euclidean	13.93%	20.65%	6.70%	9.34%	14.62%	13.53%	29.38%
	Cosine	7.76%	20.13%	3.94%	4.48%	7.98%	9.85%	20.45%
	Correlation	12.11%	21.59%	1.89%	0.69%	0.15%	0.49%	8.93%
	Spearman	7.81%	19.23%	1.89%	4.65%	8.75%	16.94%	22.71%
	Dynamic Time Wrapping	11.36%	20.22%	6.67%	8.99%	15.71%	17.22%	29.88%

Table 2.3: HTER Rates for Model-aware Approach Using Distance Measure Between PSD Feature Vectors Evaluated Against 11 Attack Types. Light Green: ≤ 10%, Dark Green: ≤ 2%, Orange: Best Case

		Thresholds						
		mean	median	mean + std	mean + 2*std	mean + 3*std	99%	max
Distance	Euclidean	11.90%	17.92%	3.35%	5.54%	4.62%	4.62%	18.06%
	Standardized Euclidean	19.92%	19.92%	3.14%	0.31%	0.00%	0.05%	0.00%
	Cosine	56.86%	61.25%	50.00%	50.00%	50.00%	50.00%	50.00%
	Correlation	56.85%	61.25%	50.00%	50.00%	50.00%	50.00%	50.00%
	Spearman	50.21%	49.85%	50.00%	50.00%	50.00%	50.00%	50.00%
	Dynamic Time Wrapping	11.43%	16.72%	3.77%	1.36%	0.42%	1.18%	4.48%

Table 2.4: HTER Rates for Model-aware Approach Using Distance Measure Between DWT Symlet (sym10) Level One Details coefficient Feature Vectors Evaluated Against 11 Attack Types. Light Green: ≤ 10%, Dark Green: ≤ 2%, Orange: Best Case

testing (97.6%).

During testing, the trained models were tested against each of the 42 type attacks separately, so the strength and weakness of the model against each of the different attacks would become clear. For testing against each attack type, the trained model was used to predict labels for the 25% of live data which was saved for testing (live class), and the attack vectors belonging to that type (adversarial class.)

		Thresholds						
		mean	median	mean + std	mean + 2*std	mean + 3*std	99%	max
Distance	Euclidean	32.77%	30.60%	36.90%	40.60%	45.07%	45.35%	45.45%
	Standardized Euclidean	17.56%	19.03%	3.09%	0.37%	3.20%	1.87%	4.55%
	Cosine	50.15%	50.35%	50.00%	50.00%	50.00%	50.00%	50.00%
	Correlation	50.30%	50.40%	50.00%	50.00%	50.00%	50.00%	50.00%
	Spearman	50.08%	50.08%	50.00%	50.00%	50.00%	50.00%	50.00%
	Dynamic Time Wrapping	21.17%	23.03%	27.95%	32.16%	38.00%	40.31%	45.27%

Table 2.5: HTER Rates for Model-aware Approach Using Distance Measure Between DWT Symlet (sym10) Level Four Approximation Coefficient Feature Vectors Evaluated Against 11 Attack Types. Light Green: ≤ 10%, Dark Green: ≤ 2%, Orange: Best Case

		Thershold Ranges						
		1-99%	5-95%	mean ± 4*std	mean ± 3*std	mean ± 2*std	mean ± std	min to max
Features	PSD	49.81%	46.35%	50.00%	48.22%	33.11%	21.69%	50.00%
	DWT-L1-cD	22.64%	20.17%	22.88%	22.80%	22.75%	22.73%	35.72%
	DWT-L4-cA	44.47%	34.85%	44.92%	44.15%	42.78%	39.16%	49.92%

Table 2.6: HTER Rates for Model Aware Approach Using Statistical Analysis of Feature Values, Evaluated Against 11 Attack Types. Green: Best Case

		Thershold Ranges						
		1-99%	5-95%	mean ± 4*std	mean ± 3*std	mean ± 2*std	mean ± std	min to max
Features	PSD	7.52%	9.30%	21.98%	13.65%	8.79%	17.99%	10.46%
	DWT-L1-cD	1.22%	5.10%	0.60%	1.14%	3.09%	11.15%	0.29%
	DWT-L4-cA	42.95%	40.99%	43.75%	42.92%	41.64%	41.16%	44.57%

Table 2.7: HTER Rates for Model Aware Approach Using Statistical Analysis of Area under Absolute Value of the Feature Vector, Evaluated Against 11 Attack Types. Light Green: ≤ 10%, Dark Green: ≤ 2%, Orange: Best Case

2.8 Evaluation

2.8.1 Model-Aware Approach Evaluation

First, the impact of machine artifacts on adversarial inputs is discussed and the distance between features derived from such signals and those from brain signals are demonstrated. Afterwards, three different methods have been suggested and evaluated against attacks.

Figures 2.3, 2.4, 2.5, 2.6 visualizes machine artifacts in four different feature domains; two wavelets and two power spectral density for the three machines considered (ANFIS, GAN, and noise addition). As seen in the figures, GAN is creating peaks in the frequency domain features at locations that are unexpected in the original signal. However, ANFIS preserves the dominant frequencies in the original signal but does not match the amplitude. For the noise addition case, changes in the range and behavior of the features are seen but depends on the amount of noise (SNR) and domain which noise was added (time or frequency). Hence, the artifact pattern of these machines are different. For proof of concept, I focused on ANFIS and GAN which are more complex machines compared to noise addition and considered impact of artifacts in the histogram of the frequency domain features (FFT in this specific case). The frequency histogram is matched with the histogram of the original signal using two distance metrics: cosine distance and dynamic time warping. The cosine distance metric can accurately distinguish between original and GAN generated signal accurately. This is because GAN introduces new dominant frequencies and changes the shape of the spectral response of the signal. However, the cosine distance metric does not discriminate between the original and ANFIS generated signal. This is because ANFIS preserve the spectral shape but fails to mimic the power density. For artifacts in frequency histogram dynamic time warping (DTW) distance is a more effective discriminator. Figure 2.7 shows this results for distance between 120 samples of original, ANFIS generated and GAN generated signals. We see that the cosine distance metric is a good discriminator between the original and GAN generated signal, but not between original and ANFIS generated signal. On the other hand DTW distance accentuates the differences between the original and ANFIS generated signals. This discussion shows for different machines there might be a need for different artifact extraction algorithms to discern fake signals from live signals. Therefore, I studied three different artifact extraction methods, three feature vectors, six distance metrics, and seven cutoff thresholds and evaluated them against attacks.

To systematically evaluate the effectiveness of utilizing machine artifacts for liveness detection, three methods were used: 1) distance metrics, 2) Distribution of Feature Values, and 3) Distribution of Area Under Feature Vector. To derive the parameters of each method only EEG signals were used, and afterwards it was tested both on EEG signals and 11 type of attacks (ANFIS-Freq, ANFIS-time, GAN, Time and Freq noise with SNRs -10,-1,1,10). The machine artifacts for these attacks are also visualized in figures 2.3, 2.4, 2.5, 2.6. Samples of length 10 seconds were used and since each of the 12 cases included 3 trails of 30 seconds for 106 subjects in total there was 954 test samples for each case. Following Chingovska et al. work on "Evaluation Methodologies for Biometric Presentation Attack Detection" (Chingovska *et al.*, 2019), we have reported Half Total Error Rate (HTER) as performance metric, which is the arithmetic mean of False Accept/Positive Rate (FAR or FPR) and False Reject/Negative Rate (FRR or FNR).

In the first method, I initially calculated the distance among EEG feature vectors using six metrics: Euclidean, Standardized Euclidean, Cosine, Pearson Correlation, Spearman, Dynamic Time Wrapping). For each case, seven cutoff thresholds from the distance values was derived: mean, median, mean plus Standard Deviation (STD), mean plus two Standard Deviation (STD), mean plus three STD, 99 percentile, and the max value. Afterwards, during testing for an incoming signal, distance between its feature vector and the EEG feature vectors in system was calculated. If for more than half of the cases, distance was less than cutoff threshold, the income would be labeled as an EEG signal, and otherwise it would be attack. The performance of these 42 combinations (six distance metrics and seven thresholds) for three different feature extraction methods (PSD, DWT Symlet (sym10) level one details coefficients, DWT Symlet (sym10) level four approximation coefficients) are reported in tables 2.3, 2.4, 2.5.

For PSD features, (table 2.3), correlation distance outperforms other measures and achieves error rate of 0.15%, and then Spearman distance (1.89%) and cosine distance (3.94%). As for the thresholds, mean plus STD and mean plus two STD provide better results in general. For DWT (sym10) level one details features (table 2.4) performance enhances and using Standardized Euclidean distance, error rate got to zero (0%). Using Dynamic Time Wrapping error was 0.42% and for Euclidean 3.35%. Finally, For DWT (sym10) level four approximation features (table 2.5, again Standardized Euclidean distance performed well and error rate was 0.37%, although other distances performed poorly.

In the second method, I estimated an distribution for EEG feature values and seven different acceptable ranges were derived: (1%-99%), (5%-95%), (mean ± 4*STD), (mean ± 3*STD), (mean ± 2*STD), (mean ± STD), (min-max). Afterwards, during testing for an incoming signal, each of its feature values were checked to see if it falls in the acceptable range or not. If for more than half of the cases, distance was less than threshold, the income would be considered as an EEG signal, and otherwise it would be attack. Table 2.6 shows the results for three different feature extraction methods (PSD, DWT Symlet (sym10) level one details coefficients, DWT Symlet (sym10) level four approximation coefficients) and seven ranges. The error rates were high and in best case for DWT level one details feature and (5%-95%) range error was 20.17%.

As the second method did not perform well, in the third method I looked into the distribution of the area under the absolute value of the feature vectors. The same seven ranges as in second method were derived, and much lower error rates was achieved (table 2.7). The best case was DWT (sym10) level one details with range of min to max with error rate of 0.29%, afterwards for PSD features error was 7.52%. The DWT (sym10) level four approximation feature performed poorly and had high error rates of around 40%. As for the ranges, the longer ones ((min-max), (1%-99%), (mean ± 4*STD)) seem to generally result in less errors.

In summary, the studied methods for model-aware approach generally performed well as one method did achieve error rate of zero (0%), another one had error of less than (0.3%), and only one method had high error rate of 20%. In model-aware approach, method parameters (i.e. threshold and ranges) are set only based on EEG signals and it was robust against attacks which were not used in any way for tuning parameters. This provides evidence that such approach has significant potential for generalization which is extremely important for defense mechanism since the attack space is so large that can never be brute-forced. So there will always be new attacks and defense techniques should be able to generalize against them without necessarily being aware of them. Finally, among the three features DWT (sym10) level one details coefficients had the best performance and PSD followed pretty closely, but DWT (sym10) level four approximation coefficients did not keep up. As for the methods, method one (distance between features) and method three (area under the absolute value of features) both were robust while method two (feature values) did not perform well. The proposed methods used lightweight signal processing techniques and did not rely on machine learning models which are more computationally expensive. Hence, such methods can be utilized locally on resource constrained devices (e.g. smartphones and embedded systems) without the need to communicate with cloud and causing large power and latency overhead.

2.8.2 Model-Agnostic Approach Evaluation

I evaluated the model-agnostic approach under the *normal* and *unseen* protocols described in section 2.7 using six machine learning models and nine feature extraction algorithms (54 combinations).

Normal Protocol

Tables 2.8 shows the performance of machine learning model and feature pairs tested against all 42 type of attacks for sample length of 10 seconds. Tables 2.9 and 2.10 show results for testing against predictive model attacks in frequency domain and time domain respectively. Tables 2.11 and 2.12 shows outcome of testing against white noise addition attacks in frequency domain (20 attack vectors) and time domain (20 attack vectors) respectively. Following Chingovska et al. work on "Evaluation Methodologies for Biometric Presentation Attack Detection" (Chingovska *et al.*, 2019), I have reported Half Total Error Rate (HTER) as performance metric, which is the arithmetic mean of False Accept/Positive Rate (FAR or FPR) and False Reject/Negative Rate (FRR or FNR).

In overall performance against all 42 attack types (Tables 2.8), the best performance was for NBC model with DWT Symlet (sym10) level one details coefficients with error rate of zero (0%). However among models, Ensemble outperformed others with most cases of error under 20% (light green boxes in table), and then Decision tree and NBC stand second. Among features DWT Symlet (sym10) level one details coefficients had best performance with low error rates of 0%, 0.04% and 0.71%. Afterwards DWT Symlet (sym10) level four approximation coefficients performs well with error rate of 0.6%.

For testing against only ANFIS generated signals in frequency domain (Table 2.9), for four combination error rate was 0% with all using th DWT features. Again the best models were NBC, Ensemble, and Decision tree and the best features were the two DWT ones. In case of testing against ANFIS generated signals in time domain (Table 2.10) the best performance was for NBC model with DWT Symlet (sym10) level one details coefficients with error rate of zero (0%). However unlike the two discussed tables, DWT Symlet (sym10) level four approximation coefficients perform poorly and instead Combine and AR performs better with error rate of 0.52% and 2.31% respectively. The same three models outperformed with best one being Ensemble.

For noise addition attacks, in Table 2.11 I tested against noise in frequency domain which achieved error rate of zero (0%) with NBC model with DWT Symlet (sym10) level one details coefficients. Ensemble was again the best model and then NBC and decision Tree and for the features the two DWT were the best same as before. In case of noise in time domain (Table 2.12), in two combinations (NBC with the two DWT features) had error rate of 0%. The same three models outperformed and beside the two DWT features, Combine and AR coefficients also performed well with errors rates of 0.92% and 0.36% respectively.

In summary, model agnostic approach using machine learning models showed robustness against attacks with having error rates of zero (0%). DWT Symlet (sym10) level one details coefficients and DWT Symlet (sym10) level four approximation coefficients would be the better choice of features, and for models Ensemble, NBC and then decision tree would be more effective. Under Normal protocol which was used in the discussed tables, models were not evaluated against new or unseen attacks.

		Feature Extraction								
		FFT	**SA**	**DWT (db1)**	**AR**	**PSD**	**Combine**	**PCA**	**DWT (sym10-L4)**	**DWT (sym10-L1)**
Model	**LDA**	24.72	34.48	45.93	25.11	41.15	9.74	50.79	41.23	37.97
	KNN	49.87	42.52	48.09	27.14	52.27	50	46.12	33.38	1.8
	SVM (RBF kernel)	46.53	46.53	46.53	17.04	49.79	46.52	50	46.52	0.71
	Decision Tree	30.64	32.95	36.16	26.88	30.61	11.98	48.92	1.73	0.04
	Ensemble	17.54	22.8	13.53	12.88	17.54	3.29	45.64	0.6	0.04
	NBC	34.94	31.45	28.88	26.04	35.6	18.06	46.47	0.63	0

Table 2.8: Half Total Error (HTER) Against All Attack Vectors. Light Green Box Shows Error Rates below 20%, Dark Green Is the Least Error Rate and Orange Boxes Show Error Rates Worse than Random Guessing (50%).

		Feature Extraction								
		FFT	**SA**	**DWT**	**AR**	**PSD**	**Combine**	**PCA**	**DWT (sym10-L4)**	**DWT (sym10-L1)**
Model	**LDA**	48.9	56.08	50.84	46.44	53.72	49.11	51.83	45.7	49.32
	KNN	78.51	60.12	64.78	46.07	74.32	80.19	45.86	33.75	0.05
	SVM (RBF kernel)	46.65	46.65	46.65	42.09	50.73	47.22	50	47.22	0.68
	Decision Tree	46.44	46.23	42.98	45.49	46.38	37.89	45.75	1.15	0
	Ensemble	37.63	36.16	20.18	34.8	37.63	41.46	46.07	0.05	0
	NBC	52.88	46.33	48.85	50.94	51.1	51	51.26	0	0

Table 2.9: Half Total Error (HTER) for Testing Against Artificially Generated Signals in Frequency Domain. Light Green Box Shows Error Rates below 20%, Dark Green Is the Least Error Rate and Orange Boxes Show Error Rates Worse than Random Guessing (50%).

Therefore, in the the next section, I evaluate how well model agnostic approach using machine learning models generalize.

Unseen Protocol

I considered four different modes of evaluation against *unseen* attack protocol (described in section 2.7):

1. **PM**: Model was trained on real signals and predictive attacks, and tested against all three types of attacks.

		Feature Extraction								
		FFT	**SA**	**DWT**	**AR**	**PSD**	**Combine**	**PCA**	**DWT (sym10-L4)**	**DWT (sym10-L1)**
Model	LDA	31.55	55.97	52.04	8.81	55.19	2.36	50.79	47.69	29.04
	KNN	65.88	61.58	50.37	12.95	67.87	65.36	49.42	42.19	5.14
	SVM (RBF kernel)	46.02	46.02	46.02	5.08	46.02	46.7	50	46.59	0.68
	Decision Tree	34.54	41.46	43.87	14.73	34.54	7.18	47.43	21.65	0.1
	Ensemble	17.14	24.16	21.96	5.61	17.14	0.52	47.8	23.06	0.1
	NBC	51.21	42.14	46.23	2.31	48.01	5.61	57.13	25.79	0

Table 2.10: Half Total Error (HTER) for Testing Against Artificially Generated Signals in Time Domain. Green Box: Least Error Rate; Orange Boxes: Error Rates Worse than Random Guessing (50%). Light Green Box Shows Error Rates below 20%, Dark Green Is the Least Error Rate and Orange Boxes Show Error Rates Worse than Random Guessing (50%).

		Feature Extraction								
		FFT	**SA**	**DWT**	**AR**	**PSD**	**Combine**	**PCA**	**DWT (sym10-L4)**	**DWT (sym10-L1)**
Model	LDA	13.05	39.13	46.52	49.05	34.11	17.51	50.89	43.56	48.37
	KNN	48.38	45.17	45.22	41.11	52.75	48.58	43.18	35.42	0.69
	SVM (RBF kernel)	46.47	46.47	46.47	28.33	50.41	46.52	50	46.52	0.74
	Decision Tree	24.02	39.12	36.48	38.71	23.99	15.78	48.55	1.35	0.05
	Ensemble	10.47	30.6	14.57	19.74	10.47	4.34	42.09	0.06	0.05
	NBC	26.03	39.5	30.12	49.72	26.23	28.54	42.62	0.04	0

Table 2.11: Half Total Error (HTER) for Testing Against Adding White Noise in the Frequency Domain to Signals. Light Green Box Shows Error Rates below 20%, Dark Green Is the Least Error Rate and Orange Boxes Show Error Rates Worse than Random Guessing (50%).

2. **TN**: Model was trained on real signals and time noise attacks, and tested against all three types of attacks.

3. **FN**: Model was trained on real signals and frequency noise attacks, and tested against all three types of attacks.

4. **TFN**: Model was trained on real signals and time and frequency attacks, and tested against all three types of attacks.

		Feature Extraction								
		FFT	**SA**	**DWT**	**AR**	**PSD**	**Combine**	**PCA**	**DWT (sym10-L4)**	**DWT (sym10-L1)**
Model	**LDA**	34.84	27.68	44.78	0.92	46.85	0.36	50.64	38.357	27.455
	KNN	49.13	38.04	50.01	12.95	49.9	49.15	48.9	30.8785	2.834
	SVM (RBF kernel)	46.61	46.61	46.61	5.08	49.32	46.48	50	46.478	0.68
	Decision Tree	36.27	25.7	35.11	14.73	36.24	7.13	49.52	1.144	0.015
	Ensemble	23.63	14.27	11.75	5.29	23.63	0.47	49.07	0.05	0.015
	NBC	42.14	22.13	25.77	2.31	43.57	6.56	49.55	0	0

Table 2.12: Half Total Error (HTER) for Testing Against Adding White Noise in the Time Domain to Signals. Light Green Box Shows Error Rates below 20%, Dark Green Is the Least Error Rate and Orange Boxes Show Error Rates Worse than Random Guessing (50%).

In normal protocol (Section 2.8.2), Ensemble model using adaptive logistic regression (LogitBoost) outperformed other models, so I decided to try another Ensemble aggregation method called XGBoost which has shown promising results in recent years (xgb, 2021). For features, we focused on feature types which showed better performance in normal protocol. Hence, we tried three DWT features (Daubechies (*db1*) level one details coefficients, Symlet *sym4* level max approximation coefficients, Symlet *sym10* level max minus two details coefficients), one AR (50th order) and one combination feature (AR 50th order and DWT Daubechies (*db1*) wavelet level one and two details coefficients), total of five different features.

Table 2.13 shows the error rates for the four evaluation modes. For samples of length 10s, training on Predictive Model (PM) attacks with AR features allows for the model to generalize with least error rate (only 0.23%). For the other evaluation cases -TN, FN, TFN- the best performance is achieved with DWT (sym4), DWT (sym10) and Combination (COMB) features respectively. Another way to look at table would be if a system is using samples of length 10s, and the designer has only access to samples of time noise attacks during training, then the best feature to detect unseen attacks would be DWT (SYM4) with 1.49% error.

Sample Length	Feature Type	Evaluation Mode			
		PM	TN	FN	TFN
10	DWT (db1)	21.80	8.55	21.80	6.43
	DWT (SYM10)	16.12	8.60	8.05	6.47
	DWT (SYM4)	6.49	1.49	8.59	2.12
	AR (50)	0.23	28.05	25.05	1.26
	COMB	1.15	28.07	25.04	1.06
5	DWT (db1)	21.08	2.93	22.26	2.43
	DWT (SYM10)	20.05	12.38	12.42	10.82
	DWT (SYM4)	6.87	1.65	12.80	2.04
	AR (50)	3.95	26.12	25.05	1.31
	COMB	7.45	26.09	25.09	1.19
3	DWT (db1)	21.05	1.84	22.51	1.74
	DWT (SYM10)	21.06	12.03	22.76	12.41
	DWT (SYM4)	7.11	1.91	13.04	2.37
	AR (50)	9.00	26.40	25.03	0.85
	COMB	4.39	26.38	25.01	0.80
1	DWT (db1)	17.73	0.75	22.86	0.91
	DWT (SYM10)	23.27	34.48	25.01	8.99
	DWT (SYM4)	8.47	3.53	10.26	3.71
	AR (50)	18.41	25.71	25.00	0.71
	COMB	18.55	25.69	25.00	0.78
	Average	12.71	15.13	20.13	3.42

Table 2.13: Half Total Error (HTER) for Testing Against Unseen Attacks for Four Different Evaluation Modes: Trained on (PM: Predictive Model Attacks, TN: Time Noise Attacks, FN: Frequency Noise Attacks; TFN: Time and Noise Attacks) and Tested Against All Three Type of Attacks. For Each Sample Length (10, 5, 3, and 1 Second), Dark Green Box Shows the Least Error among the 20 Cases (Five Features & Four Modes), and the Light Green Box Shows the Least Error for Each Mode.

Furthermore to improve system usability, I explored using shorter sample lengths of 5,3 and 1 seconds beside samples of 10 seconds (as in normal protocol). Impact of decreasing sample length on performance was not uniform and depended on both the feature type and the evaluation mode; in around half of the cases it increased the error, in quarter of them error stayed in the same range, and in the other quarter it even decreased the error. For example, in case of DWT (db1) feature and TN mode, error rates decreases significantly (8.55%, 2.93%, 1.84% and 0.75% for sample length of 10, 5, 3 and 1 seconds respectively).

In the last row, the average error rate for each column is reported and TFN with average error rate of 3.42% has the best performance. In other words, training the model on noise addition attacks on time and frequency, would be the best approach for system designer to generalize well against unseen PM attacks and also the noise addition attacks themselves. Putting TFN mode aside (as it trains on two attack types) then PM with average error rate of 12.71% outperforms TN (15.13%) and FN (20.13%). However, in the best case scenario (light green boxes) TN does better compared to PM for samples of 5s (1.65% vs 3.95%), 3s (1.84% vs 4.39%) and 1s (0.75% vs 8.47%). Only for best case of 10s samples TN loses to PM (1.49% vs 0.23%). So while on average case training with PM mode is better, in best case scenario TN mode performs better.

In summary, to have best performance against unseen attacks, system should be trained on either predictive attacks or time noise attacks for best generalization. Training on frequency noise attacks should be avoided as it generally has poor performance.

2.9 Conclusion

In this chapter, I discussed how the assumed *intrinsic liveness property* for brain signals in literature does not hold and brain signals are vulnerable to pretension attacks and require liveness detection methods. To help bridge the gap, I formulated the brain liveness problem and proposed two solution approache: *model-aware* with focus on detecting machine artifacts in fake signals, and *model-agnostic*) with focus on finding effective computational features in brain signals. Under each approach several methods were suggested and evaluated against attacks (43 types). Both synthetic (e.g. generative and predictive models) and manipulative (e.g. noise addition) attacks were considered and the methods were robust with several cases of zero (0%) error rate. Furthermore, the methods generalized well against unseen attacks with is an significant factor in quality of defense mechanism as the search space for attacks cannot be brute-forced and new attacks can and will appear.

Chapter 3

BIOMETRIC LIVENESS DETECTION

In this chapter, I will first discuss Biometric Liveness Detection (BLD) problem, and its challenges, system model, solution approaches and real-world examples of presentation attacks including the emerging *DeepFake* attacks. Afterwards, I will take the first steps toward building evaluation criteria for comparing liveness detection methods, and evaluate the current state-of-the-art with it, and also provide a taxonomy of the related works. Finally, I argue that liveness detection should shift away from biological features more toward cognitive capabilities and limitations of humans compared to machines and provide examples of novel methods for liveness detection.

3.1 Background

I will discuss definitions, problem statement and system model to provide a background on the Biometric Liveness Detection (BLD) problem

3.1.1 Definitions

It was not until 2016, when International Organization of Standardization (ISO) and the International Electro-technical Commission (IEC) in a manuscript titled *Biometric presentation attack detection* (ISO, 2016) provided formal definitions, which we will use as base in this work, and customize them for our work.

- **Liveness:** “quality or state of being alive, made evident by anatomical characteristics, involuntary reactions or physiological functions, or voluntary reactions or subjects behavior.” (ISO, 2016).

- **Liveness Detection:** "measurement and analysis of anatomical characteristics or involuntary or voluntary reactions, in order to determine if a biometric sample is being captured from a living subject present at the point of capture." (ISO, 2016).

- **Presentation Attack:** "presentation to the biometric data capture subsystem with the goal of interfering with the operation of the biometric system." (ISO, 2016). Examples of adversarial inputs mimicking valid human traits are replaying/generating voice commands or brain/heart signals, gummy finger with valid user fingerprints on it, displaying user pictures (printed or on screen) to a face recognition system, etc.

Liveness detection goal is to verify two properties: *live property:* input has been sensed from a live human subject, *timely property:* input has being sensed at the current point in time and is not a replay of an previous recording.

Presentation attacks has also been refereed to as *spoofing attacks* in literature, and similarly liveness detection has been termed as *presentation attack detection*. I will be using these terms interchangeably in the rest of the paper, and the same goes with *human input*, *input sample* and *trait*.

3.1.2 Problem Statement and System Model

The Biometric Liveness Detection (BLD) problem can be formulated as follow; *Considering a system equipped with biometric access control module which receives human traits as inputs, how should the system distinguish between live input and adversarial input?*

Formally, the problem can be defined as follow (Equation 3.1):
Considering set A containing samples from a source S, decision function G is needed

which on system input I outputs 1 if the input belongs to the source. For more detailed discussion on formal problem definition, check section 2.3.

$$\begin{aligned} &A = \{a_i | a_i \in S\} \\ &G(I, A) = \begin{cases} 1, & I \in S \\ 0, & otherwise \end{cases} \end{aligned} \tag{3.1}$$

Figure 3.1, shows a generic attack model, where an adversary provides adversarial input (i.e. old/fake trait) through biometric sensor which goes through matching process to be compared with the stored trait (more details in section 3.7). Based on the result of matching, a decision will be made to either reject the trait or accept it which will lead to providing access to the main system. For the matching process component the typical goal is authentication; to match the incoming trait with the claimed user's signature in system. However, authentication cannot prevent adversaries to gain access through replay of past genuine inputs or artificially generated inputs which mimic the original one (e.g. gummy finger). Reason is such adversarial inputs do exactly or closely match the user signature and from the authentication point of view they are valid. system. Hence, there is a need for liveness detection methods to prevent old or fake date to be utilized for gaining unauthorized access to system.

3.2 Solution Approaches

BLD solutions in the literature primarily try to exploit biological characteristics of traits and there are two main approaches for it: 1) *Challenge and Response* (CR), 2) *Baseline Monitoring* (BM). In Challenge and Response (CR) approach, the subject is challenged and it is expected to provide a response. For instance, the subject might be asked to do an specific physical movement (e.g. blink) or perform a task (e.g.

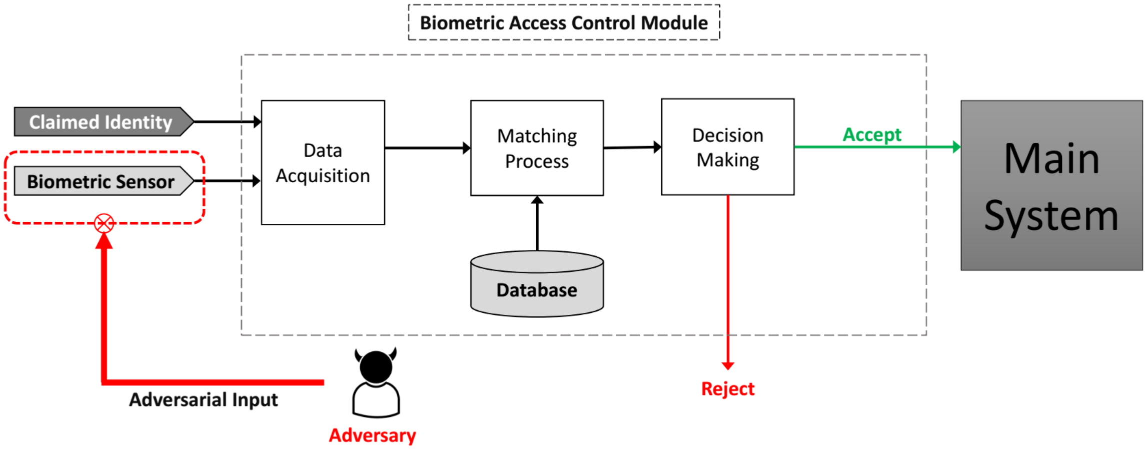


Figure 3.1: Attack Model for Biometric Access Control Module

choose images containing cats). The underlying assumption is that the response to the challenge requires a human individual at point of input and it can not be faked by a machine. In Baseline Monitoring (BM) approach, while subject is not performing any specific task, her trait(s) is sensed and based on them BLD is performed. There are three sub-categories in the BM approach; a) *Software-based Solution* (SS), b) *New Coherent Modality* (NCM), and c) *New Incoherent Modality* (NIM). In Software-based Solution (SS) category, liveness detection is performed on the same trait which is used for the authentication purpose. These methods attempt to derive new set of features from the input trait to enable detecting liveness from them. In New Coherent Modality (NCM) category, a new trait dependent on the same source is sensed. For example, in case of fingerprint, skin conductivity is added for liveness detection; this new modality is sensing a new trait but from the same finger. In New Incoherent Modality (NIM) category, a new trait independent of the main source is sensed. For instance, adding EEG monitoring (i.e. brain signals) for Liveness detection in a face recognition based authentication system.

These approaches have their own shortcomings. For CR methods, since the challenges are known, the human adversary can monitor the environment and provide the appropriate response at time of the challenge to a machine at the scene. The SS methods are vulnerable to stolen input traits since system is designed in a manner which accepts user inputs and has no way of distinguishing between the stolen and the original input. In case of NCM/NIM methods, these new modality inputs can be generated and get accepted by the security system (Xu *et al.*, 2016; Sadeghi *et al.*, 2017; Maiorana *et al.*, 2013; Raghavendra *et al.*, 2017; Duc and Minh, 2009; Eberz *et al.*, 2017).

Using bio-electrical signals such as EEG (brain) and ECG (heart) is generally suggested as the current best practice due to an claimed *intrinsic* liveness characteristic for them. This intrinsic feature is based on a assumption that these signals can only be captured from a live human body (Matthew, 2016; Barra, 2016). However, recent works (including one from me) have shown the feasibility of generating artificial bio-electrical signals, that get accepted by current biometric systems (Sadeghi *et al.*, 2017; Maiorana *et al.*, 2013).

Beside these shortcomings, there can be another potential serious blow for these approaches. Biofrabrication, which is the science of manufacturing human organ and tissue, can fundamentally undermine the effectiveness of such biological based approaches to BLD. There are already successful works in generating mini-heart (Ma *et al.*, 2015), mini-kidney (kid, 2013), mini-lung (lun, 2015) and female genitals (Raya-Rivera *et al.*, 2014) which demonstrates the possibility of organ manufacturing. Therefore, moving toward new approaches are an urgent need for this domain.

A new direction would be to shift BLD methods toward intelligence and cognitive capabilities of human and focus on brain abilities which machines still haven't achieved. The current state of art in BLD, focuses on the trait and tries to determine if it had originated from a live human being. The suggested shift would target the objective goal of BLD; instead of focusing on the trait for liveness inference, the goal should be to check if the *entity* providing the input sample is alive. For this goal, the attention should be on cognitive abilities of human rather than its biologic characteristics. Some examples of this new approaches are mentioned below and discussed comprehensively in section 3.11.

- **Emotional Reaction**: checking how the user reacts to an emotional stimuli e.g. a scary movie or a joke.

- **Human Errors**: Exploiting relative visual perception and optical illusions challenges that lead to human error, but machines can answer error-free.

- **Rapid Human Learning**: Teaching and asking at the same time. Human can learn fast while machine training is time consuming.

- **System History**: Asking about past interactions with the system.

In literature, the input trait has been used for both authentication and liveness detection task. Going beyond that two scenarios regarding the sample usage can be considered. First, adding physiological and behavioral context to the environment and afterward detecting the expected effect of the added context on the sample as a liveness sign. For example, in the case of fingerprint, the subject can be asked to twist her finger at a certain time, or in case of face, have the subject rise her eyebrows. Second, a new set of inputs from the subject will be captured with the sole intent of liveness detection. An example would be to ask the subject, to read aloud some CAPTCHA texts.

3.3 Challenges

While Liveness Detection (LD) seems like a trivial task, there are many obstacles in this way. I will cover the main challenges in accurate and robust detection of liveness.

- **Ground Truth** The first and perhaps the most important challenge is that there is no concrete ground truth for liveness in an adversarial environment. This means there is no inherent sign or characteristic in an input trait that guarantees its liveness. This implies no feature can be derived from the input sample that definitely ensures it has originated from a live human being. This issue is to some extent related to not having tangible measures for defining liveness.

- **Machine Interrogator** Liveness detection is performed by a machine and therefore its performance is limited to the machine capabilities. Although machines have significantly become smarter in recent years –even winning human champions in games such as Chess and Go–, they still suffer from fundamental shortcomings such as lack of consciousness and common sense. Machines can only execute the code written by a human programmer and have no idea of the task they are performing either if it is calculating sale tax or forecasting the weather. So machines can only run automated methods of liveness detection and lack understanding of life concept.

- **Matching Method** There does not exist any deterministic algorithm for liveness detection of a sample input unlike password checking. This is because of the nature of liveness detection problem that lacks any concrete ground truth as discussed above. All the proposed methods are non-deterministic (e.g. probabilistic, heuristics, etc.); usually working based on comparing the input with some previous sample data and then making decision based on the matching score. The current state of art relies on statistical and machine learning methods for this purpose.

- **Performance and Robustness Trade-off** In non-deterministic methods used for liveness detection, the matching score between the current input and the previous sample data is compared with a threshold for decision making. If it is above the threshold, it is accepted as live and otherwise it is rejected as not being live. As it can be seen, this threshold plays an important role in performance of liveness detection by determining to which degree the input should be similar to the previous samples. On one hand, as threshold is set to higher values, less number of fake samples will get accepted (false accepts) but at the same time, false rejection of valid samples will become higher. On the other hand, if the threshold is set to a lower value to decrease the false rejects, it will allow more fake sample to get accepted by system. There is no value for threshold that will satisfy the two objectives of having zero false reject and zero false accept rate.

- **Evaluation** Liveness methods should not only be evaluated in laboratory settings and abstract from real-world trade-offs such as cost (e.g. financial, power, latency), usability, security level and privacy. Instead they should be evaluated in the deployed environment with respect to them mentioned trade-offs. Also the evaluations can not solely rely on experimental/numerical approaches, due to issues such as high computation complexity, lengthy/infeasible computation time, results being limited to specific datasets, lack of scalability to general case, and finally providing estimated results with no guarantees. Therefore, analytical approaches are preferred which do not suffer from those shortcomings and can provides guarantees. Also lack of suitable metrics for comparing different methods is an important issue which should also be addressed to allow better design decisions.

- **Usability** Similar to other fields of security, usability is usually neglected in designing liveness detection methods. There are instances of more robust and secure methods (implemented in research labs) compared to those used in real-world systems. However, since they might suffer from usability shortcomings they would not be deployed in practice and less robust but more usable techniques will be utilized.

- **Sample Collection and Generation** Biometric samples can be collected in various ways without a subject's consent and knowledge. For example, fingerprints are available on most of the surfaces touched by the subject, or subject's face image can easily be found on Internet or can be captured by a camera even from far distances. Moreover, samples can be artificially generated using models of traits (Sadeghi *et al.*, 2017; Maiorana *et al.*, 2013). This collected or generated samples can later on be presented to system to gain access. The problem is how should the liveness detection method reject these samples while they belong to the authentic subject.

3.4 Surveys on Biometric Liveness Detection

In the last two decades, numerous works have been published on BLD methods based on different traits, however number of survey works are not as many. In this section, I will focus on the survey literature of the recent years and how my provided taxonomy differs. Surveys in literature are either very narrow and deep (i.e. discussing methods for one specific trait in a systematic manner) or broad and shallow (i.e.

discussing methods for few traits without systematic analysis and categorization). I aim to be broad and deep, meaning that we will systematically analyze and provide taxonomy of methods for wide range of the traits in literature.

Most surveys have focused on techniques for only a specific trait. Ramachandra and Busch (2017) did a survey of BLD techniques for face recognition, Marasco and Ross (2015) for fingerprint-based recognition, Czajka and Bowyer (2018) for iris-based recognition, and Wu *et al.* (2015a) for voice-based recognition. Our work instead will cover these traits and other ones in literature such as palm print, handwriting and thermal scan in order to generate a comprehensive overview of techniques in the BLD domain which allows for three advantages. First, it will familiarize readers in each sub-domain (e.g. a specific trait) with methods used in other sub-domains which can potentially be exploited for their problem. Second, it will assist in design of BLD methods for multi-modal biometric systems (i.e. systems using more than one trait) which requires knowledge of methods in all and each of the traits. Finally, it can facilitate creation of new techniques from fusion of the state-of-art ones.

There are few surveys that cover a wider range of traits or look into general issues in BLD methods. Akhtar *et al.* (2015) mostly discusses the research opportunities and challenges in BLD and provides an high level and brief discussion on the general ideas of BLD methods using face, finger and iris without categorizing them. Hadid *et al.* (2015) starts with explaining spoofing attacks (another term for presentation attacks) and biometric systems' vulnerability to them, and then describes an evaluation method for BLD techniques. However, the evaluation method (accuracy and error rates based on false accept and false reject rates) is same as the one used for authentication systems throughout years which only captures the performance and not any other factors (e.g. cost, usability, etc.). Matthew and Anderson (2014b) discusses appropriate characteristics of a trait to be useful for in BLD methods. The discussed characteristics closely overlap with the ones for suitable trait for authentication goal (e.g. universality, uniqueness, collectability, acceptability, etc.). This thesis instead provides a systematic taxonomy of BLD methods which is not limited to the common traits (i.e. fingerprint, face, iris, voice) and categorizes the methods based on different characteristics. Furthermore, a multi-factor evaluation criteria is proposed and methods are evaluated and scored based on it, which had not been done in the literature.

In summary, this chapter will help toward bridging three gaps in BLD domain. First, it will provide a comprehensive survey of BLD methods, covering traits-based (e.g. biological) and also non-trait-based (i.e. cognitive) techniques. Second, design of a multi-factor evaluation criteria and evaluating the surveyed works based on it. Finally, proposing new classes of Liveness detection methods for biometric authentication systems.

3.5 Taxonomy of Liveness Detection Methods

Table 3.2 provides a taxonomy of BLD in literature plus my proposed methods (marked by *). The methods are divided into two main categories based on stage in which liveness detection happens: 1) detection *during data acquisition*, and 2) detection *after data acquisition*. In the first category, beside the primary traits which are collected for the main task (e.g. authentication), other inputs are collected from subject for the specific task of liveness detection. In the second category, liveness is to be determined only based on the primary traits collected from subject and there is no specific input for liveness task.

3.5.1 During Data Acquisition Methods

In these methods, nature of the new sensed input can be *physiological* which focuses on biological features, or *psychological* with focus on cognitive and intellectual aspects, and in both case it might require a new sensing hardware.

physiological input

Reddy *et al.* (2008) used pulse oximetry for fingerprint to detect liveness in which saturation of oxygen in hemoglobin and also heart pulse is measured. Derakhshani *et al.* (2003); Parthasaradhi *et al.* (2005) exploited the point that perspiration temporally changes moisture patterns on live finger while dead/fake fingers to do not show such changes. Skin distortion is also used as liveness feature, in which subject was asked to apply pressure on fingerprint scanner and rotate the finger (Antonelli *et al.*, 2006). Drahansky (2008) experimented skin electric resistance and skin temperature for fingers as liveness features. Authors concluded that skin resistance and temperature are not the best candidate since the range of their measurements for live humans is broad enough to include fake fingers too. Czajka (2015), used pupillary response to light (changes in size of eye pupil) for liveness detection in iris.

Psychological Input

In these group of methods, context is added and process of data acquisition is personalized so that an entity with intellectual/cognitive capabilities can successfully complete it. The literature has mostly neglected these set of techniques, although ISO/IEC standard (ISO, 2016) provides examples such a requesting head node, closing left eye and random order of fingers. Here, I will provide some other examples of such methods which are discussed with more details in section 3.11.

Different individuals can express different feelings toward same stimuli and machines are incapable of expressing feeling at least for now. Therefore system can display fearful/humorous images and detect its effect in subject's face, voice or pulse. Humans are capable of fast paced learning while machines learning process is long and complex, so liveness detector can teach some material in real-time and ask about it. For certain pattern recognition tasks machines are error-free while humans do make mistakes and errors. This can utilized by displaying a graphical illusion which fools human but not machine (basically an inverse CAPTCHA). Another approach would be to ask questions regarding the previous interactions with system (login times, login locations, task performed, etc.) about which only the valid user is knowledgeable. It's noteworthy that first three of these methods are only effective against machine adversary and fail against human adversary, while the last one is effective against both.

3.5.2 After Data Acquisition Methods

These methods can be studied in two separate sets: 1) looking for new type of features in the collected data, and 2) designing new feature extraction algorithms.

Psychological New Features

These methods attempts to discover and exploit new characteristics of live human traits which is absent in synthesized or lifeless inputs. Shiota *et al.* (2015) used pop noise (voice distortion in microphone caused by noise of breathing) as a sign of liveness for voice traits. Natural eye blinking has been considered as a liveness feature for face against photo attacks which are motionless (Li, 2008; Sun *et al.*, 2007). In another works, progressive eyelid tracking (Ghosh and Negi, 2016) and eye movements Jee *et al.* (2006); Komogortsev *et al.* (2015) are used as a sign of liveness against mechanical replicas.

Psychological Software Features

These methods search for new software-based feature extraction algorithms to increase the separability between live and lifeless biometric traits. Ghiani *et al.* (2012) used a texture classification algorithm (i.e. local phase quantization) to detect differences in spectrum characteristics of fingerprint images caused by loss of information during production of adversarial input. Rattani and Ross (2014) suggest using a secondary classifier to detect finger spoofs made of unseen materials (not included in training set) to help adaptively enhance the primary liveness classifier performance against new spoof materials. Galbally *et al.* (2012b) integrates different group of features (e.g. focus, motion, and occlusion) to train classifiers capable of distinguishing between live iris and high quality images.

3.6 Real World Presentation Attacks Examples

Adversaries are successfully using presentation attacks for hacking into real world devices such as smartphones. Step-by-step tutorials are available on Internet which allow common user to preform presentation attack with minimum cost and effort. Therefore, such attacks are not limited to researchers with exceptional knowledge of the field or state-entities with excessive resources, and can be launched fairly easy by numerous entities. This poses serious concern for security and privacy of users since we are observing a trend of migration from traditional authentication methods (e.g. password, pin) to biometric authentication methods in almost all domains (e.g. medical, smartphones, broader control, etc.). In this section, I will discuss several effective presentation attacks against biometrics systems and also investigate the recent video-based attacks termed as *DeepFake*, where seemingly realistic but fake videos of users is crafted behaving by the adversary's desire.

I have gathered some examples of both presentation attacks from literature performed in laboratory settings and from online news/tutorials performed in real world, as shown in Table 3.1.

Matsumoto *et al.* (2002) made fake fingerprints from gelatin and their samples were accepted by 11 different fingerprint sensors (optical and capacitive) with at least 67% success rate. In a similar work Galbally *et al.* (2011) used silicone to craft fake fingers which were accepted by three type of sensors (optical, capacitive and thermal) using two matching algorithms (minutiae-based and ridge feature-based). Both works consider two cases: 1) with user cooperation: user presses its finger on a gelatin mold, and 2) without user cooperation: latent fingerprint obtained from a surface (e.g. glass, CD) touched by the user.

Tome *et al.* (2014) used images of palm vein printed with commercial printers to bypass the system with 65% success rate in pool of 50 subjects. In a similar work, Ruiz-Albacete *et al.* (2008) printed fake iris images with commercial printers, and reached success rate of 71% - 99% in bypassing iris-based recognition system.

Xu *et al.* (2016) developed a 3D facial model from user's social media images in virtual reality with success rate of 58% against 5 commercial face recognition systems (industry level) equipped with both liveness and motion detection modules. The 3D models had texture and capable of performing facial expressions such as smiling and raising eyebrows. Raghavendra *et al.* (2017) crafted printed images using commercial printers that could easily fool face recognition systems (96% success rate) that use multi-spectral cameras with 7 bands ranging between 425- 930nm. Authors also tested four liveness detection methods in literature and showed they have medium to low performance with respect to this type of attack. Duc and Minh (2009) used printed images to fool face recognition systems on three laptop brands (i.e. Asus, Lenovo, Toshiba).

Carlini *et al.* (2016) created audio signals that were difficult to comprehend for humans, but contained hidden commands easily recognized by voice recognition systems. Authors test their crafted signals against Samsung Galaxy S4 and Apple Iphone 6, which on average were recognized by phones in 60% of the attempts and only understood by human listener in 46% of the attempts. Eberz *et al.* (2017) synthesized artificial ECG signals using 3 different methods; hardware-based Arbitrary Waveform Generator (AWG), software AWG working on sound card, replay of ECG signals with audio player. Testing on Nymi Band, authors achieved 43% - 81% success rate.

DeepFake

Most of the attacks listed in Table 3.1, can easily be detected by an human observer rendering them ineffective. However, new type of attacks are surfacing that can (at least) in the first look fool human observer too. Fake but realistic videos have gone viral in the last few years showing public figures saying things they never have said (e.g. president Obama (net, 2018d), Facebook founder Mark Zuckerberg (net, 2019d)) or performing actions never done (celebrity/revenge pornography (net, 2018b)). These type of videos are termed as *DeepFake* since they are artificially crafted with deep neutral networks. Creating DeepFake videos does not require special knowledge or extensive resources, in contrary there are ready to use software and tutorials (FakeApp (net, 2019b), DeepNude (net, 2019f), DeepFaceLab (net, 2018a), MyFakeApp (net, 2019e)) even from research community (Nvidia vid2vid (Wang *et al.*, 2018; net, 2018c)) that allows a common user to synthesize such videos on home machines in fairly short time (e.g. 24-72 hours). The essential material required is images/videos of the target individual and with increase in the amount of image/video fed to deep neutral networks in training phase, higher quality videos can be crafted. In case of public figures there is an abundance of their high quality image/video publicly available online which results in realistic videos in return.

The ease of creating DeepFake videos for non-expert users with limited resources, allows this type of presentation attack to pose severe threat especially considering its capability in fooling human observer. Beside the usage of DeepFake for propaganda and revenge goals, these videos can potentially be used to exploit current state of the art in face-based recognition systems. In a more complex scenario, an adversary can create such videos in real time and use it to impersonate entities even in case of holding a dialogue with another individual (e.g. interview, video calls). The adversary itself can talk with the victim, but on the fly a crafted video showing target's face consistent with her voice will be broadcast. This attack can be even more automated by using a Natural Language Processing (NLP) system to understand victim voice and provide responses without the need of an human adversary.

In traditional presentation attacks, human observer would easily detect the fake trait and could assist the liveness detection systems. These more advanced attacks, might create an urgent need for liveness detection mechanisms for human users in near future. The question then would be how to design liveness detection systems that can assist individuals in pointing out artificial materials (e.g. video, image, audio, text).

With the emergence of DeepFake videos in the recent years, countermeasures have been proposed in literature. Güera and Delp (2018) trained Recurrent Neutral Networks (RNN) to detect anomalies in DeepFake videos based on intra-frame and temporal inconsistencies. They reached 97% success rate on a dataset of 600 videos, which half was DeepFake. Authors exploited three flaws in such videos as follows. First, different camera angles, lighting condition and video codecs in videos of target affect the quality of produced video and result in visual inconsistencies. Second, only the face region of the video is being manipulated and swapped which creates boundary effects with rest of the frame. Third, manipulation is performed individually on each frame, and therefore results in temporal inconsistencies between frames which creates artifacts such as flickering effect on face area. Li and Lyu (2018) proposed detecting resolution inconsistencies caused by affine transformations (i.e. scaling, rotation, shearing) used in face wrapping in DeepFake videos (usually low resolution ones). They tested their method against two dataset and in best case reached Area Under Curve (AUC) performance of 93% to 99%. In other works lack of natural eye blinking (Li *et al.*, 2018) and inconsistent head pose (Yang *et al.*, 2019) has been suggested for detecting DeepFake videos.

These mentioned works tested their methods on datasets which had been crafted using the few publicly available software applications and possibly performed by common users. Therefore, the results can not necessarily be generalized to DeepFake videos made by other means or by expert users.

3.7 System and Threat Models

In this section, I describe a generic system model and three threat model for biometric authentication systems.

3.7.1 Biometric Authentication System

In this chapter, I focused on Biometric Authentication Systems (BAS) that use machine learning models to match the input with user's signature in system, and provide a potential match as output. As seen in Figure 3.2, the generic Biometric Authentication Systems (BAS) model has the following components:

Table 3.1: Successful Presentation Attacks against Biometric Authentication Systems.

Biometrics	**Attacks**
Fingerprint	Gummy finger (Matsumoto *et al.*, 2002)
	latent fingerprint on surface (Galbally *et al.*, 2011)
Vein	Spoofing sample collection (Tome *et al.*, 2014)
	Wolf Attack (Une *et al.*, 2007)
Face (Image/Video)	Building virtual models from public photos (Xu *et al.*, 2016)
	Printed Image (Raghavendra *et al.*, 2017; Duc and Minh, 2009)
Voice	Hidden voice commands (Carlini *et al.*, 2016)
	Audio Reply (Alegre *et al.*, 2014)
Iris	Fake images (Ruiz-Albacete *et al.*, 2008)
Heart Signal (ECG)	transformed signals (Eberz *et al.*, 2017)

Data Acquisition

Different types of sensors have been designed to record biometric traits such as fingerprints, iris, and brain Electroencephalography (EEG) signals. Sensors record analog data from human body and send it to data acquisition module. In data acquisition phase, recorded analog data is converted into digital format. Unlike deterministic security systems that use passwords, biometric data is not same for each person due to non-uniform sensor setup, different environmental conditions, measurement errors, etc. In an ideal case, the collected biometric data should have low intra-subject and high inter-subject variability. In general, biometric raw data has high dimensionality and comparing data samples from various persons is a computationally intensive task. Extracting features with lower dimensions can help the BAS to process the data more efficiently.

Liveness Detection Methods	**During Data Acquisition**	**Physiological (new metrics)**	(1) Pulse oximeter (Reddy *et al.*, 2008)	Fingerprint
			(2) Blood pressure reading (Rogmann and Krieg, 2015)	Fingerprint
			(3) Blood volume pulse probing (Liu *et al.*, 2016)	General
			(4) Requesting physical inputs e.g. weight, hair, urine, etc.*	General
			(5) Perspiration in fingerprinting (Derakhshani *et al.*, 2003; Parthasaradhi *et al.*, 2005)	Fingerprint
			(6) Ultrasonic images (Gu *et al.*, 2016)	Fingerprint
			(7) Opto-electronic fingerprinting (Kiss *et al.*, 2001)	Fingerprint
			(8) Skin distortion scans (Antonelli *et al.*, 2006)	Fingerprint
			(9) Skin electric resistance and skin temperature (Drahansky, 2008)	Fingerprint
			(10) Infrared and ultraviolet light images (Matthew and Anderson, 2014a)	General
			(11) Thermal scans (Matthew and Anderson, 2014a)	General
			(12) Pupillary response to light (Czajka, 2015)	Iris
			(13) Reflectance analysis (Kose and Dugelay, 2013)	Face image
			(14) Eye closing request (Rogmann and Krieg, 2015)	General
			(15) Head turning request (Rogmann and Krieg, 2015)	General
			(16) Fusion of ECG and fingerprint (Komeili *et al.*, 2018)	Fingerprint
			(17) Adding throat microphones (Sahidullah *et al.*, 2018)	Voice
			(18) Taking photo with and without flash (Chan *et al.*, 2017)	Face (image)
			(19) Plethysmographic Signals (Krishnan *et al.*, 2018)	Finger vein image
		Psychological	(20) Emotional reaction such as fearing, laughing, etc.*	General
			(21) Teaching and asking in real-time*	General
			(22) Exploiting human errors and mistakes*	General
			(23) Questions about the last activities in the system*	General
			(24) Request of different fingers in random order (Rogmann and Krieg, 2015)	Fingerprint
			(25) Questions out of field of expertise*	General
			(26) Philosophical questions e.g. the concept of time*	General
			(27) Interactive facial expression (Ming *et al.*, 2018)	Face (image/video)
			(28) Attention-based filtering mechanism (Lai *et al.*, 2019)	Voice
	After Data Acquisition	**Physiological (new features)**	(29) Pop noise in voice caused by breath (Shiota *et al.*, 2015)	Voice
			(30) Progressive eyelid tracking (Ghosh and Negi, 2016)	Face image
			(31) Natural eye blinking (Li, 2008; Sun *et al.*, 2007)	General
			(32) Eye movements (Jee *et al.*, 2006; Komogortsev *et al.*, 2015)	Face image
			(33) Natural muscle movements while speaking (Matthew and Anderson, 2014a)	General
			(34) Unique vibration pattern of both human vocal cord and throat (Shang *et al.*, 2018)	Voice
			(35) Gaze alignment to a moving stimulus (Alsufyani *et al.*, 2018)	Face (video)
		Physiological (Software-based feature extraction)	(36) Software-based fingerprint detection (Ghiani *et al.*, 2016)	Fingerprint
			(37) Using quality related fingerprints features (Galbally *et al.*, 2012a)	Fingerprint
			(38) Characteristics of blood flow (Lapsley *et al.*, 1998)	Iris
			(39) Local phase quantization (Ghiani *et al.*, 2012)	Fingerprint
			(40) Automatic adaptation to new spoof materials (Rattani and Ross, 2014)	Fingerprint
			(41) Wavelet-based detection (Moon *et al.*, 2005)	Fingerprint
			(42) Image power spectrum (Coli *et al.*, 2007)	Fingerprint
			(43) Using thin-plate spline distortion model (Zhang *et al.*, 2007)	Fingerprint
			(44) Algorithmic-based counter measures (Tome *et al.*, 2015)	Vein image
			(45) Texture and 3D structure analysis (Lin *et al.*, 2016)	Face (image/video)
			(46) Sparse low rank bilinear discriminative model (Tan *et al.*, 2010)	Face (image/video)
			(47) Locally uniform comparison image descriptor (Ziegler *et al.*, 2012)	Face (image/video)
			(48) Fourier spectra analysis (Li *et al.*, 2004)	Face (image/video)
			(49) Optical flow and structure tensor analysis (Kollreider *et al.*, 2005)	Face (image/video)
			(50) Using local descriptors (Gragnaniello *et al.*, 2015)	Iris
			(51) Demodulation by complex-valued wavelets (Daugman, 2003)	Iris
			(52) Using quality related features (Galbally *et al.*, 2012b)	Iris
			(53) Morphology analysis of physiological signals*	General
			(54) Correlation among physiological signals*	General
			(55) Fusion of physiological signals*	General
			(56) Extracting features based on Eulerian video magnification (Aydoğdu *et al.*, 2018)	Palmprint (video)
			(57) Alignment of audio and video using dynamic time wrapping (Aides *et al.*, 2018)	Audiovisual

Table 3.2: The Taxonomy of Liveness Detection Methods (* Indicates the Proposed Methods in This Work). The Last Column Lists the Corresponding Trait.

Feature Extraction

After acquiring raw data using biometric sensors, some mathematical and statistical methods are employed to extract data features (Equation 3.2). A feature extractor function, $F(.)$, maps input data from an n-dimensional domain $\mathcal{D}$ to an m-dimensional domain $\mathcal{F}$. Given input data, X, derives a feature vector (FV).

$$\begin{aligned} &F(.) : \mathcal{D}^n \to \mathcal{F}^m \\ &X = \{x_1, x_2 \dots x_n\} \\ &F(X) = FV = \{fv_1, fv_2, \dots fv_m\} \end{aligned} \tag{3.2}$$

In the registration phase, a set of input data are obtained from the subject (e.g. Sub_i) and the corresponding extracted feature vectors are stored in a database and categorized as a *subject class* (e.g. C_{Sub_i}). In the future authentication attempts by that subject, the feature vector is derived from incoming input and than matched with stored vectors in class C_{Sub_i} through classification process. Indeed, feature extraction filters randomness and encodes unique characteristics of a given input data (X), that can be used to distinguish subjects from each other according to their biometric traits. For example, some common feature extraction methods used in brain Electroencephalography (EEG) based authentication systems include: Power Spectral Density (PSD), Fast Fourier Transform (FFT), Discrete Wavelet Transform (DWT), and AutoRegressive (AR) coefficient (Nicolas-Alonso and Gomez-Gil, 2012).

Feature extractors can be *reversible* (Equation 3.2). There exists a function $F^{-1}(.)$ such that given a feature vector, FV, in domain $\mathcal{F}$ outputs input, X, in domain $\mathcal{D}$. Note that the function $F(.)$ can be a many to one function. In such a scenario, F^{-1} can not be mathematically determined since different input data result in the same exact feature vector. For example, two inputs X and Y can have the same mean and standard deviation features. Hence, from the derived mean and standard deviation features function F^{-1} cannot determine if the original input had been X or Y. In such cases, feature extractors are *irreversible*, since there is no unique mapping from features domain to input domain. Another *irreversible* case is when given the feature vector, FV, one cannot find a an input X such that $F(X) = FV$ as with cryptographic hash functions (He *et al.*, 2009).

$$\begin{aligned} &F^{-1}(.) : \mathcal{F}^m \rightarrow \mathcal{D}^n \\ &F^{-1}(FV) = X \\ &F^{-1}(\{fv_1, fv_2, \dots fv_m\}) = \{x_1, x_2 \dots x_n\} \end{aligned} \tag{3.3}$$

Classification and Decision Making

Machine learning techniques are widely used for classification in security systems such as Naïve Bayes Classifier (NBC), Support Vector Machine (SVM), and Neural Networks (NNs) (Del Pozo-Banos *et al.*, 2014).

Equation 3.4 formulates such techniques where given a feature vector (FV), a claimed identity (Sub_i), and parameter set (P) a binary machine $M(.)$ computes to what extent feature vector matches to subject class C_Sub_i space and the rest of space termed as world class, C_W. In decision making stage, based on comparing the two computed level of matching with the spaces, output (or label) will be one it is decided that feature vector belongs to the claimed subject (Sub_i) and zero otherwise (belongs to the world class, C_W).

$$\begin{aligned} P &= \{p_1, p_2 \dots p_k\} \\ M(FV, Sub_i, P) &= \begin{cases} 1, & FV \in C_{Sub_i} \\ 0, & FV \notin C_{Sub_i} \equiv FV \in C_W \end{cases} \end{aligned} \tag{3.4}$$

There can be multi-class machines which compare the features with more than two classes but binary machine are typically utilized for authentication purposes. Multi-class machines are useful for *identification* task were the goal is to determine if the input data belongs to some subject in database such as fingerprint matching in forensics applications.

The machine uses the parameter set, P, to decide whether a given data falls in class C_{Sub_i} or not. This parameter set is derived through a training mechanism. The training mechanism uses a set of data (training set) T_D and their true labels ($\{0, 1\}$), and uses a series of algorithms depending upon the machine to determine the parameters P such that classifying the training set (T_D) has the least error compared to their true labels. The quality of future classification and decision making depends on how well the machine has been trained and to what extent it can generalize.

3.7.2 Attack Scenarios and Threat Model

The authentication process can be done in a *supervised* setting (e.g. military facilities, forensics, or maybe any situation that the access port to the biometric system is under video surveillance or human monitoring) or an *unsupervised* setting (e.g. biometrics for mobile phones login). Obviously, in the second setting, the adversary has extremely more freedom in exploiting the system than the first one, making it a more complex problem. I chose the second case as threat model.

There exist a large body of work on BAS and its vulnerabilities has been discussed (Jain *et al.*, 2006). Figure 3.2 indicates critical attack points in the given system model. The different attack points require access to various types of information for the adversary. More information about the system is potentially equivalent to less effort by adversary to break the system. In this research, I focused on presentation attacks, which are related to the earliest entry point of the system model

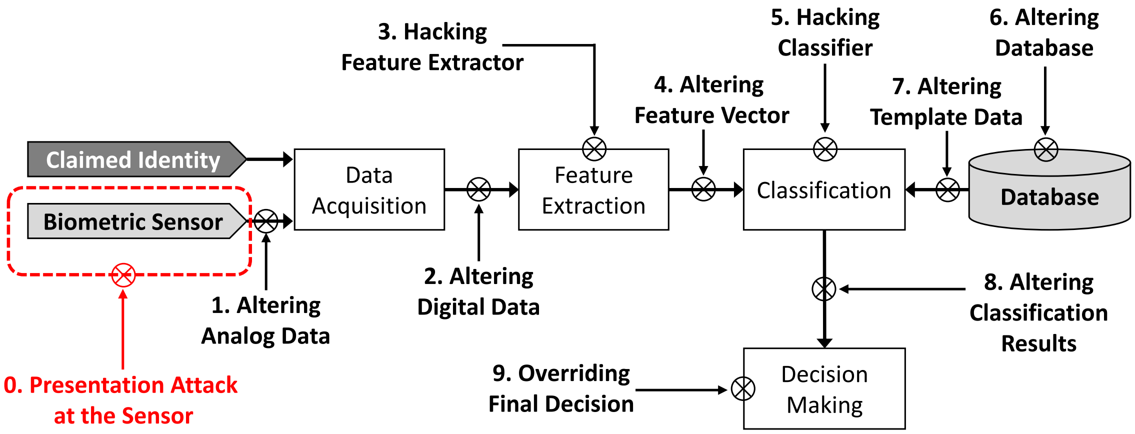


Figure 3.2: Possible attack points in a BAS.

(point 0 in Figure 3.2). The main aim of a presentation attacker is to present a fake biometric sample to the sensor which would get accepted by the classification and decision making unit as a valid user. Gummy fingertip with genuine subject fingerprint or high quality printed face image are some examples of presentation attack. The success of these attacks depends on available tools for the adversary and level of her knowledge about the system. In my threat model, adversary can be a human equipped with natural intelligence, or a humanoid machine with some level of consciousness using artificial intelligence such as an Internet bot. Below definitions for different intelligence levels is provided.

Definition 3.7.1. *Human Intelligence* (HI) is the ability of perceiving surrounding, making decision accordingly, and perform an action based on the decision to enhance the chance of survival (Wang and Wang, 2006). HI is not perfect and continuously evolving.

Definition 3.7.2. *Machine Intelligence* (MI) is an imitation of natural intelligence using computer programs. Although, MI has not generally bypassed the HI so far, but in some tasks has better performance than HI.

Definition 3.7.3. *Humanoid Machine* is a standalone autonomous intelligent agent, rather than a passive machine intelligence program, who knows the attacking task and decides how to attack the BAS based on MI.

Definition 3.7.4. *Tools* are any available devices that assist the adversary to increase the chance of successful attack, such as computer software/hardware, composites, 3D printer, etc.

Threat Model

I discussed the threat model with particular reference to presentation attacks against BAS. Based on the different roles of the adversary, threat model covers three main attack scenarios as seen in Figure 3.3. In all the attack scenarios, adversary can only interact with the system by providing biometric raw data as input. Hence, even if the adversary crafts a biometric feature vector, it has to be converted back to raw biometric data before the adversary can present it to system. In such scenarios, I assume that the BAS only uses reversible feature extractors. Since if irreversible features are crafted, reproducing it to raw input data might not be feasible. The three attack scenarios are as follows:

1. **Human:** There is a human adversary with access to tools for presentation attack as seen in Figure 3.3A. In this scenario, the adversary is equipped with HI and tools.

2. **Machine:** There is humanoid machine in role of an adversary with access to tools to fool the system to be recognized as a live subject (Figure 3.3B). In this case, the adversary has MI and tools.

3. **Hybrid:** In this scenario, human and machine cooperate, and use tools to break

into the BAS as a live genuine subject as seen in Figure 3.3C. In this attack, the adversary has both HI and MI plus tools.

In the threat model (Figure 3.3), adversary can have different level of knowledge from the BAS as described below. In general as adversary's knowledge about system increases, the vulnerability of the BAS against presentation attack also increases.

- **Data Acquisition:** including knowledge such as data sampling precision, range of data, and data sample.
- **Database:** access to the subjects data samples stored in the database during registration phase.
- **Feature Extraction:** knowledge about feature extraction algorithm, range of feature data, and feature calculation precision.
- **Classification:** knowing about classification algorithm mechanism, matching scores, and classification performance.
- **Decision Making:** knowing about decision making mechanism and thresholds.

The main objective goal of the adversary is to present a raw input data such that the resulting feature vector can a) bypass the liveness detection test, and b) get classified as subject class.

3.8 Evaluation Criteria

In this section, I discuss evaluation criteria for BLD methods with regard to three main factors (i.e. usability, cost, performance) and later on in Section 3.10, evaluate the methods proposed in this work and surveyed ones based on them. These criteria help with comparing different liveness detection methods, and also help with choosing the most appropriate method based on system requirements. Moreover, these criteria should be considered at design time in order to have effective LD methods in the end.

3.8.1 Usability

While Usability is usually neglected in research works, it is one of the most important criteria which will determine the system overall success in real usage. Many of the proposed methods in the security domain provide high level of robustness but because of their low usability subjects are not willing to use which makes them useless in practice. Below, I will discuss the most important usability criteria for liveness detection methods.

- **Task length:** The length of time the subject should spend on interacting with system for the purpose of liveness detection. This time can range from being almost instant (e.g. order of milliseconds, ms) to several seconds/minutes. Shorter tasks are preferred.

- **Task effort:** The amount of effort the subject needs to devote while interacting with the system for the purpose of liveness detection. The subject effort can range from completely effortless (e.g. thermal mapping) to high effort (e.g. reading a CAPTCHA text). Methods with less effort is preferred.

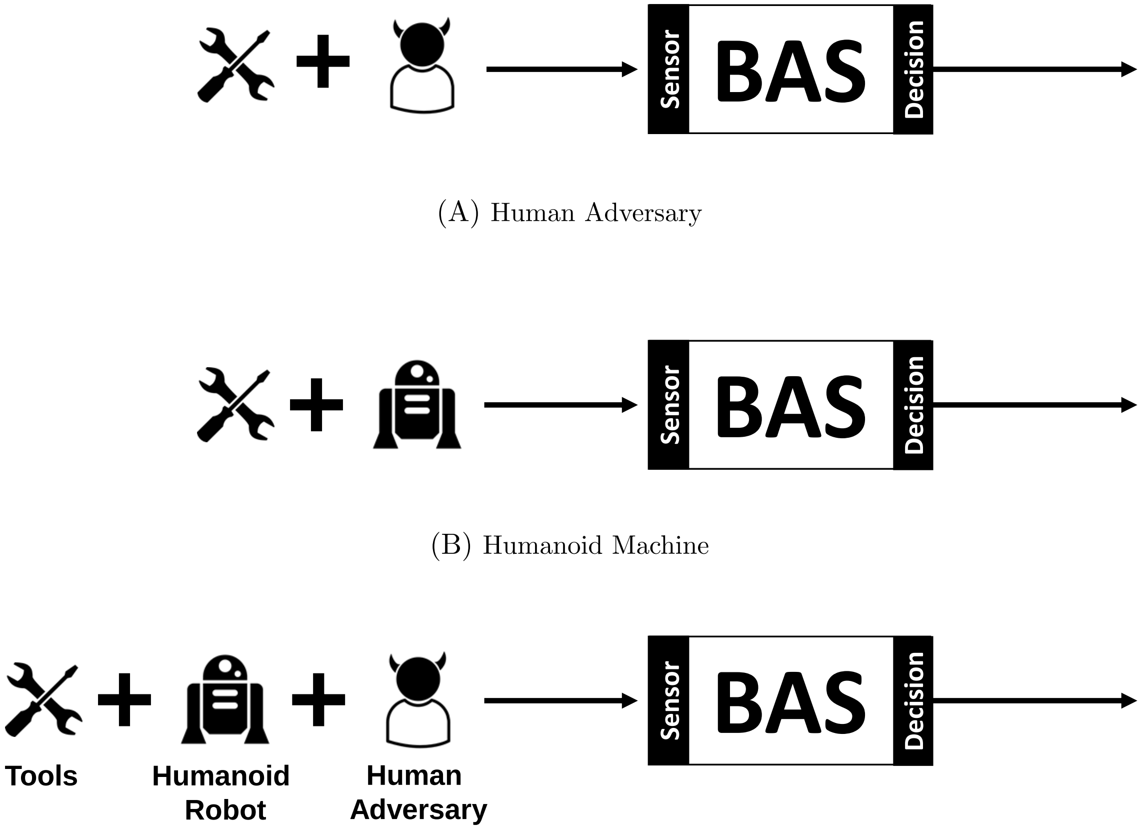
- **Extra Task** Some methods ask the subject to perform a single task for both the authentication and liveness, while other methods have separate tasks for authentication and liveness. An example of the first group is skin conductivity checking in fingerprinting systems where the subject only need scan her fingers; subject's fingerprint is used for the authentication purpose and skin conductivity for liveness. In the second group, a face recognition method where the subject is asked to blink at a certain time is an example of the second group; face is used for the authentication purpose and blinking for liveness. A single task method is preferred.

- **Sensor Interaction Level** Some methods require the subject to wear a sensor (e.g. heart and brains monitoring), while in some other the sensor is not wearable and direct interaction is needed (e.g. fingerprinting, key stroke), and finally some require indirect interaction (e.g face and voice recognition). Methods with less sensor interaction is preferred.

3.8.2 Cost

Methods are generally developed in research laboratory settings where their cost (not only financial, but also others required conditions for real-world usage) is usually not a priority. However, in time of deployment cost is one of the major parameters. Below, I have discussed the different cost criteria associated with liveness detection methods.

- **Financial:** The immediate financial cost for deploying a method depends on its equipment costs which include sensor and computing unit (capable of handling the computation) costs. For instance, ultrasonic imaging cameras are more expensive than regular ones, and fingerprinting computation can be done in a simple embedded machine, while image/video processing need more powerful (and therefore more expensive) machines. Equipment also have long-term financial cost for maintenance and electricity usage. Moreover indirect costs of the amount of space occupied by equipment should be considered. In corporation buildings available spaces is an asset and in mobile devices space is extremely



(A) Human Adversary

(B) Humanoid Machine

(C) Cooperation of Human and Machine

Figure 3.3: Three Main Attack Scenarios for Presentation Attacks Against Biometric Authentication Systems.

limited (therefore large sensors are unfavorable). A method with less financial costs is preferred.

- **Computation:** Methods have different computation complexity required for data processing and decision making based on the nature of the input trait and algorithms used. For instance, processing images is more complicated than voice, and machine learning techniques are more complex than signal processing ones. As the computation complexity increases, more powerful machines would be needed to handle the computation. These methods are to be used in wide range of environments and machines (cloud servers to embedded/mobile systems). Therefore computation complexity should be appropriate for this varying range. Simpler methods (if it does not degrade the performance) is preferred.

- **Latency:** The amount of time the method needs to decide about the liveness of the incoming input. Latency depends on the computation complexity of the method and computation power of the machine. Lower latency is preferred.

- **Power:** The average amount of energy per second the machine requires to execute the method. Power usage is especially important for mobile devices that rely on battery, and methods which would drain the battery fast are unfavorable. Even on machines with permanent power supply (e.g. desktop or server machines), power usage is not neglected cause it results in electrical bill. Power usage depends on the computation complexity of the method and technological properties of the machine executing the algorithms. Lower power usage is preferred.

- **Extra sensor:** Considering an already operating biometric authentication system, the liveness detection method might exploit the same input used for authentication or might require an extra sensor for a new input for liveness purpose. For instance, in a face recognition system, liveness detection method might use the same face image, or choose to perform thermal imaging which then requires a thermal camera. Methods which do not require a new sensor are preferred.

3.8.3 Performance

Liveness detection methods are designed to prevent presentation attacks. The method performance can be measured by its effectiveness against different presentation attacks. These attacks can be performed in various ways based on the assumed attack model (system vulnerable points and also adversary knowledge about system mechanism and data) and also assumed attack scenarios (adversary type, expertise and capabilities). It is not feasible to consider all attack models and attack scenarios for determining method performance, therefore in this work we chose to use the same attack model defined by International Organization of Standardization (ISO) for presentation attacks (ISO, 2016), as seen in Figure 3.2. Also for attack scenarios, we consider three cases (Figure 3.3) that summarizes the main possible scenarios. In these scenarios, there are three entities; human, machine, tools. We consider two state for each of them; normal and intelligent for human and machine, elementary and advanced for tools. Therefore, there would be four situations for scenario A and B (e.g. advanced tools and normal human for scenario A or elementary tools and intelligent machine for scenario B), and 8 situations for scenario C (Figure 3.3). The methods will be evaluated against these 16 attack scenarios.

A liveness detection method can have two types of error; *False Accept* and *False Reject*. In case of False Accept (FA), an adversarial input get verified as live input and bypasses the liveness detection method. Such error has security implications as it can lead to unauthorized access by adversary. In False Reject (FR), a live input gets verified as adversarial and therefore gets rejected by system. This error does not cause security implications but instead usability ones as a valid user cannot get access to systems and needs to try again. Depending on application, one error type might be more critical than the other and such requirement should be considered during

LD design phase. Half Total Error (HTER) which is defined as arithmetic mean of False Accept (FA) and False Reject (FR) rates, aggregates the two error types and is widely used in BLD field (Chingovska *et al.*, 2019).

3.9 Input Trait Criteria

The traits(s) which are being used as input to liveness detection methods, should satisfy specific criteria to be suitable for this purpose. In Biometric Authentication System (BAS), six criteria are considered for a sample trait Jain *et al.* (2004), however for liveness domain, not all of these criteria are vital and only four out of these six ar important (I did not discuss performance here for two reasons, first its not related to input nature, and second we had a discussion on performance in Section 3.10). I will first discuss the four crucial criteria, and afterward go over the other two and explain why they are not that much significant for LD task.

But before that and as a side note, there might be a need for another type of input trait for BAS. While liveness detection can prevent presentation attacks, it is ineffective against coercion attacks Matthew and Anderson (2016), in which a valid subject is forced by an adversary to perform the authentication. An example of coercion attacks is when in a bank robbery the banker is threatened by gun to unlock a safe equipped with biometric authentication system. Clearly, since a valid live subject is providing input it will easily bypass authentication and liveness checks, hence a different set of techniques and inputs are required to detect and prevent coercion attacks. Therefore, in BAS up to three set of input traits might be required; one for access control (core goal of system), another for liveness checking (to prevent presentation attacks), and a last one for coercion checking (to prevent coerced attacks). However, coercion attacks are outside the scope of this thesis andis only mentioned to provide a more comprehensive picture of BAS requirements.

3.9.1 Universality

A input sample used for liveness detection should have high level of universality meaning it should be common between all or almost all individuals. Fingerprint and face are two decent input trait; although there are individuals who do not have fingerprints due to genetic problems or being disabled with hand amputations. In case of face, even if it gets burnt or disfigured, it can possibly still be used for recognition. Voice is also another option, however its not suitable for deaf and speech-impaired individuals. Inputs which check intelligence level of the entity providing it, can be more appropriate because it depends on the brain which is an inseparable organ of a live human. However, intellectually disable individuals will have issues providing these type of inputs. The thermal map of human body might be an ideal input because of its universality among live human beings.

3.9.2 Collectability

The ease of collecting an input sample from subjects is an important factor in the liveness detection process. Collectability improves as systems move from input samples that require wearable sensors (e.g. brain and heart monitoring) to those which need direct interaction with sensor (e.g. fingerprinting, key stroke), to finally samples with indirect sensor interaction (e.g face and voice recognition). The input samples which can be collected in a seamless manner might seem favorable but on the other hand rises issues with regards to ethics and subject privacy. It is clear that samples with invasive sensing (e.g. blood testing) are not preferred.

3.9.3 Acceptability

The input sample should be acceptable by law and cultural values of the society. Otherwise, the system will not be used if subjects feel offended or uncomfortable because of the input trait itself or its collection manner. Moreover, the sensing input sample should not be associated with any health issues, risks and even rumors. For instance, individuals might think that through brain monitoring their thoughts can be recorded, which will result in subjects not accepting the system.

3.9.4 Circumvention

Ideally the input trait should poses characteristics that makes it impossible to be generated artificially or collected without subjects cooperation. Although, such trait is not at our disposable, and any human trait can be learned and then be crafted. Also most of the inputs, beside those that require wearable sensors, can be collected without subject knowledge. So the inputs that are harder to be circumvented are preferred. For instance, fingerprint can be collected from the surfaces touched by subject, face and iris image can be captured from distance, voice can be recorded by devices in close proximity of subject, and DNA can be obtained form saliva/hair samples. On the other hand, collecting heart/brain signals are much more difficult and also highly challenging to be collected without subject knowledge.

3.9.5 Uniqueness

For liveness detection there is no need that the input sample should necessarily be unique to each individual because the purpose here is not specific subject authentication, but checking if the entity providing input is a live human being. Although if a sample has this property, it might help with liveness detection performance. An example of a trait which is not distinctive to each subject but proper for liveness would be the thermal map of human body.

3.9.6 Permanency

Same as with uniqueness property, input sample do not need to be stable through time and under different conditions. Once again because the target here is to detect liveness, not to do authentication. Most of the inputs are not stable under long period of time, although some change more frequently than others. For instance, while bio-electrical signals are time-dependent in nature, the iris is considered permanent to a large extent. On the other hand, face is permanent under periods of few days or weeks, but not few month/year.

3.10 Methods Evaluation and Trade-Offs

Evaluating methods and cross-comparing them against each other will provide insights into their strengths and drawbacks. This type of analysis will come in handy at system design time to better chose the appropriate method. I have performed comprehensive evaluation of the methods with regards to the three main categories of criteria: usability (four criteria), cost (five criteria) and performance.

Methods' performance have been thoroughly analyzed under 16 different attack scenarios, as seen in Table 3.3. In this table, the columns are categorized in three main groups based on the attributes of the adversary: 1) (tools, human), 2) (tools, machine), and 3) (tools, human, machine). In each group of columns, the strength of the adversary increases from left to right, which can lead to decay in the performance of the liveness detection method under study. The performance index in the last column indicates the sum of qualitative performances for all 16 threat model, which shows in overall, how a given liveness detection method operates in different

Table 3.3: Performance of liveness methods against different threat models.

	Tools: *Elementary* Human: *Normal*	Tools: *Advanced* Human: *Normal*	Tools: *Elementary* Human: *Intelligent*	Tools: *Advanced* Human: *Intelligent*	Tools: *Elementary* Machine: *Normal*	Tools: *Advanced* Machine: *Normal*	Tools: *Elementary* Machine: *Intelligent*	Tools: *Advanced* Machine: *Intelligent*	Tools: *Elementary* Human: *Normal* Machine: *Normal*	Tools: *Advanced* Human: *Normal* Machine: *Normal*	Tools: *Elementary* Human: *Intelligent* Machine: *Normal*	Tools: *Advanced* Human: *Intelligent* Machine: *Normal*	Tools: *Elementary* Human: *Normal* Machine: *Intelligent*	Tools: *Advanced* Human: *Normal* Machine: *Intelligent*	Tools: *Elementary* Human: *Intelligent* Machine: *Intelligent*	Tools: *Advanced* Human: *Intelligent* Machine: *Intelligent*	**Performance Index (Max = 16.0)**
(1) Pulse oximeter	◐	○	◐	○	◐	○	◐	○	◐	○	◐	○	◐	○	◐	○	**4.0**
(2) Blood pressure reading	◐	○	◐	○	◐	○	◐	○	◐	○	◐	○	◐	○	◐	○	**4.0**
(3) Blood volume pulse probing	◐	○	◐	○	◐	○	◐	○	◐	○	◐	○	◐	○	◐	○	**4.0**
(4) Requesting physical inputs	○	○	○	○	●	◐	●	○	○	○	○	○	○	○	○	○	**2.5**
(5) Perspiration in fingerprinting	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	**12.0**
(6) Ultrasonic images	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	**12.0**
(7) Opto-electronic fingerprinting	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	**12.0**
(8) Skin distortion scans	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	**12.0**
(9) Skin resistance	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	**12.0**
(10) Infrared and ultraviolet light images	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	**12.0**
(11) Thermal scans	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	**12.0**
(12) Pupillary response to light	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	**12.0**
(13) Reflectance analysis	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	**12.0**
(14) Eye closing request	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	**12.0**
(15) Head turning request	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	**12.0**
(16) Fusion of ECG and fingerprint	◐	○	◐	○	◐	○	◐	○	◐	○	◐	○	◐	○	◐	○	**4.0**
(17) Throat microphone	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	**12.0**
(18) Taking photo with and without flash	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	**12.0**
(19) Plethysmographic signals	◐	○	◐	○	◐	○	◐	○	◐	○	◐	○	◐	○	◐	○	**4.0**
(20) Emotional reaction	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	**12.0**
(21) Teaching and asking	◐	◐	◐	◐	●	●	●	●	●	●	◐	◐	●	◐	◐	◐	**11.5**
(22) Human error and mistakes	◐	◐	◐	◐	●	●	●	●	●	●	◐	◐	●	◐	◐	◐	**11.5**
(23) Last interactions with the system	●	●	●	◐	●	●	●	◐	●	●	●	◐	●	◐	●	◐	**13.5**
(24) Request of different fingers in random order	◐	○	◐	○	◐	○	◐	○	◐	○	◐	○	◐	○	◐	○	**4.0**
(25) Knowledge test	◐	◐	◐	◐	●	●	●	●	●	●	◐	◐	●	◐	◐	◐	**11.5**
(26) Philosophical questions	◐	◐	◐	◐	●	●	◐	◐	◐	◐	◐	◐	◐	◐	◐	◐	**9.0**
(27) Interactive facial expression	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	**12.0**
(28) Attention-based filtering	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	●	◐	**12.0**
(29) Pop noise in voice caused by breath	●	◐	◐	◐	●	◐	◐	◐	●	◐	◐	◐	◐	◐	◐	◐	**9.5**
(30) Progressive eyelid tracking	●	●	●	◐	●	●	●	◐	●	●	●	◐	●	◐	●	◐	**13.5**
(31) Natural eye blinking	●	◐	◐	◐	●	◐	◐	◐	●	◐	◐	◐	◐	◐	◐	◐	**9.5**
(32) Eye movements	●	◐	◐	◐	●	◐	◐	◐	●	◐	◐	◐	◐	◐	◐	◐	**9.5**
(33) Natural muscle movements while speaking	●	◐	◐	◐	●	◐	◐	◐	●	◐	◐	◐	◐	◐	◐	◐	**9.5**
(34) Vocal cord and throat vibration	●	◐	◐	◐	●	◐	◐	◐	●	◐	◐	◐	◐	◐	◐	◐	**9.5**
(35) Gaze alignment	●	◐	◐	◐	●	◐	◐	◐	●	◐	◐	◐	◐	◐	◐	◐	**9.5**
(36) Software-based fingerprint detection	●	●	●	◐	●	●	●	◐	●	●	●	◐	●	◐	●	◐	**13.5**
(37) Using quality related fingerprints features	●	●	●	◐	●	●	●	◐	●	●	●	◐	●	◐	●	◐	**13.5**
(38) Characteristics of blood flow	●	●	●	◐	●	●	●	◐	●	●	●	◐	●	◐	●	◐	**13.5**
(39) Local phase quantization	●	●	●	◐	●	●	●	◐	●	●	●	◐	●	◐	●	◐	**13.5**
(40) Automatic adaptation to new spoof materials	●	●	●	◐	●	●	●	◐	●	●	●	◐	●	◐	●	◐	**13.5**
(41) Wavelet-based detection	●	●	●	◐	●	●	●	◐	●	●	●	◐	●	◐	●	◐	**13.5**
(42) Image power spectrum	●	●	●	◐	●	●	●	◐	●	●	●	◐	●	◐	●	◐	**13.5**
(43) Using thin-plate spline distortion model	●	●	●	◐	●	●	●	◐	●	●	●	◐	●	◐	●	◐	**13.5**
(44) Algorithmic-based counter measures	●	●	●	◐	●	●	●	◐	●	●	●	◐	●	◐	●	◐	**13.5**
(45) Texture and 3D structure analysis	●	●	●	◐	●	●	●	◐	●	●	●	◐	●	◐	●	◐	**13.5**
(46) Sparse low rank bilinear discriminative model	●	●	●	◐	●	●	●	◐	●	●	●	◐	●	◐	●	◐	**13.5**
(47) Locally uniform comparison image descriptor	●	●	●	◐	●	●	●	◐	●	●	●	◐	●	◐	●	◐	**13.5**
(48) Fourier spectra analysis	●	●	●	◐	●	●	●	◐	●	●	●	◐	●	◐	●	◐	**13.5**
(49) Optical flow and structure tensor analysis	●	●	●	◐	●	●	●	◐	●	●	●	◐	●	◐	●	◐	**13.5**
(50) Using local descriptors	●	●	●	◐	●	●	●	◐	●	●	●	◐	●	◐	●	◐	**13.5**
(51) Demodulation by complex-valued wavelets	●	●	●	◐	●	●	●	◐	●	●	●	◐	●	◐	●	◐	**13.5**
(52) Using quality related features	●	●	●	◐	●	●	●	◐	●	●	●	◐	●	◐	●	◐	**13.5**
(53) Morphology analysis of physiological signals	●	●	●	◐	●	●	●	◐	●	●	●	◐	●	◐	●	◐	**13.5**
(54) Correlation of physiological signals	●	●	●	◐	●	●	●	◐	●	●	●	◐	●	◐	●	◐	**13.5**
(55) Fusion of physiological signals	●	●	●	◐	●	●	●	◐	●	●	●	◐	●	◐	●	◐	**13.5**
(56) Eulerian video magnification	●	●	●	◐	●	●	●	◐	●	●	●	◐	●	◐	●	◐	**13.5**
(57) Alignment of audio and video	●	●	●	◐	●	●	●	◐	●	●	●	◐	●	◐	●	◐	**13.5**

● = **High performance** ◐ = **Medium performance** ○ = **Low performance** attack scenarios. For example in method number 16 (i.e. fusion of heart (ECG) and fingerprint), the detection performance for a normal adversary with elementary tools is high. However, an adversary with advanced tools (e.g. with capability of tampering the heart (ECG) sensor) can undermine the detection performance. As another example, the method number 17 (i.e. using throat microphone) is vulnerable when facing an intelligent machine equipped with advanced tools (e.g. with capability of learning the behavior of human vocal system and tampering the microphone). However, usage of throat microphone can block the attacks launched by normal machines equipped with elementary tools. Finally, method number 57 (i.e. alignment of audio and video) is a software-based method, which shows high performance facing several threat models. For instance, an intelligent adversary with elementary tool and a normal machine cannot break this method (e.g. it is not capable of tampering microphones and cameras for successful attack). However, an intelligent adversary with advanced tools and an intelligent machine can recreate the exact behavior of the user and undermine the performance.

The overall evaluation results based on usability, cost and performance is available in Table 3.4. In this table, there are three main factors for evaluating the liveness detection methods: 1) usability, 2) cost, and 3) performance. The overall score, in the last column aggregates the scores from all three factors. For example, method number 16 (i.e. fusion of ECG and fingerprint), since additional body sensors are added to the system, the usability is low and there will be some delay in obtaining the results from the side channel. Adding sensor also increases the cost including the price of the new hardware and computational cost (processing the extra collected data). However, the cost is usually paid off by reaching higher performance in detection. As another example, in method number 32 (i.e. eye movements), no additional sensors is required for liveness detection, which means higher usability and lower cost.

However, the collected data needs extra processing to extract new features, which will increase the cost to some extent. Based on the efficiency of the processing module the performance may vary. Finally, in software-based methods such as Fourier spectral analysis (number 48), the usability is high. Also, the cost is high due to extra in-depth processing of the data. However, in-depth analysis of the features typically leads to high detection performance.

Since this evaluation is done in qualitative manner, my opinion has effects on its outcome, however I tried my best to perform a fair evaluation. Without doubt, future research needs to pay special attention in designing more quantitative evaluation criteria and methodologies to allow for more precise comparison. Although, it is noteworthy that to best of my knowledge this thesis is the first step toward any kind of comprehensive evaluation criteria which go beyond performance metrics.

Table 3.4, also shows the trade-offs between these three criteria for each method. As it can be observed, there is no method that satisfies all three criteria and joint optimality of them can not be achieved. In the following sections, I discuss the two-way and three-way trade-offs in liveness detection domain.

3.10.1 Performance vs. Usability

Liveness detection performance is generally constrained by usability aspects. More robust methods can be designed but since they would require long and high effort tasks, they would not be usable. Currently many of the proposed methods are also not in use because the usability criteria had not been optimized in them. The important point is not to forget the role of subject in the system. Subject is the primary entity interacting with system, and the more the subject feels comfortable, the more system would succeed in its goals. Designing the system with a subject-centric perspective will help to overcome the trade-off between performance and usability to some extent.

3.10.2 Performance vs. Cost

On one hand, performance can be improved if more accurate sensors and more complex algorithms are leveraged. Also using various sensors will increase the method robustness against attacks. But on the other hand, these upgrades in sensors would result in financial and space cost rise. More complex algorithms and the more powerful computing units needed for handling them, will increase the latency and power costs. Generally higher performance levels are correlated with higher costs.

3.10.3 Performance vs. Usability vs. Cost

In order to improve performance to the highest possible levels, it would be needed to have multi-factor sensing based on different interactive tasks combined with advanced algorithms. This level of performance would negatively impact usability and cost since it would require multiple sensors and subjects high effort interaction with them, and also increase in computation, latency and power costs due to the usage of complex algorithms. These systems might be useful only in applications where performance is the number one priority (e.g. military facilities).

3.11 Solutions and Approaches

As mentioned in Section 3.4, liveness solutions can be applied in two stages; 1) during data acquisition, and 2) after data acquisition. For the first stage, I suggest *psychological/memory*-based interaction with the subject to collect more information to enhance the chance of liveness detection. But, in the second stage, the opportunity of collecting more information from the subject has passed, so deriving new features from available information can be helpful. In this case, multi-biometric approaches including, *correlation* among traits, biometrics *fusion*, and knowledge-based tests are

Table 3.4: Evaluation of liveness methods.

	Usability	*Task length*	*Task effort*	*Extra task*	*Sensor interaction level*	**Cost**	*Financial*	*Computation*	*Latency*	*Power*	*Extra sensor*	**Performance**	**Overall Score**
(1) Pulse oximeter	◐	◐	◐	○	●	●	●	◐	◐	●	●	○	○
(2) Blood pressure reading	○	●	◐	○	●	●	●	◐	◐	●	●	○	○
(3) Blood volume pulse probing	○	●	◐	○	●	●	●	◐	◐	●	●	○	○
(4) Requesting physical inputs	○	◐	●	●	◐	●	●	◐	◐	●	●	○	○
(5) Perspiration in fingerprinting	●	◐	○	○	◐	●	●	◐	◐	●	●	◐	◐
(6) Ultrasonic images	●	○	○	○	○	●	●	◐	◐	●	◐	◐	◐
(7) Opto-electronic fingerprinting	●	○	○	○	○	●	●	◐	◐	●	◐	◐	◐
(8) Skin distortion scans	●	○	○	○	○	●	●	●	●	●	○	◐	◐
(9) Skin resistance	●	○	○	○	○	●	●	●	●	●	●	◐	◐
(10) Infrared and ultraviolet light images	●	○	○	○	○	●	●	●	●	●	●	◐	◐
(11) Thermal scans	●	○	○	○	○	●	●	●	●	●	●	◐	◐
(12) Pupillary response to light	●	◐	○	◐	○	●	●	●	●	●	○	◐	◐
(13) Reflectance analysis	○	◐	●	●	◐	◐	◐	◐	◐	◐	◐	◐	○
(14) Eye closing request	○	◐	◐	●	◐	◐	◐	◐	◐	◐	◐	◐	○
(15) Head turning request	○	◐	◐	●	◐	◐	◐	◐	◐	◐	◐	◐	○
(16) Fusion of ECG and fingerprint	○	●	○	○	●	●	●	◐	◐	◐	●	●	◐
(17) Throat microphone	●	○	○	○	◐	◐	◐	◐	◐	◐	●	●	●
(18) Taking photo with and without flash	●	○	○	○	○	◐	○	◐	◐	◐	○	●	●
(19) Plethysmographic signals	◐	○	○	○	◐	◐	◐	◐	◐	◐	●	◐	●
(20) Emotional reaction	◐	◐	◐	●	○	●	◐	●	●	●	◐	◐	○
(21) Teaching and asking	○	●	●	●	◐	●	◐	●	●	●	●	◐	○
(22) Human errors and mistakes	◐	◐	◐	●	○	●	◐	◐	◐	●	◐	◐	○
(23) Last interactions with the system	○	◐	●	●	◐	○	○	◐	◐	○	◐	●	●
(24) Request of different fingers in random order	●	○	◐	●	○	○	○	○	◐	○	○	○	●
(25) Knowledge test	○	◐	◐	●	◐	◐	●	◐	◐	○	◐	◐	○
(26) Philosophical questions	○	◐	●	●	◐	○	○	●	◐	○	◐	◐	○
(27) Interactive facial expression	○	◐	◐	◐	○	◐	○	●	◐	●	○	●	●
(28) Attention-based filtering	○	○	○	○	○	◐	○	●	◐	●	○	●	●
(29) Pop noise in voice caused by breath	●	○	○	○	○	○	○	◐	◐	◐	○	◐	◐
(30) Progressive eyelid tracking	●	◐	○	○	○	◐	◐	◐	◐	◐	◐	●	●
(31) Natural eye blinking	●	◐	○	○	○	◐	◐	◐	◐	◐	◐	◐	●
(32) Eye movements	●	◐	○	○	○	◐	◐	◐	◐	◐	◐	◐	●
(33) Natural muscle movements while speaking	●	◐	○	○	○	◐	◐	◐	◐	◐	◐	◐	●
(34) Vocal cord and throat vibration	○	○	○	○	○	◐	○	◐	◐	◐	○	●	●
(35) Gaze alignment	◐	◐	◐	◐	○	◐	◐	◐	◐	◐	○	●	●
(36) Software-based fingerprint detection	●	○	○	○	○	●	○	●	●	●	◐	●	●
(37) Using quality related fingerprints features	●	○	○	○	○	●	○	●	●	●	◐	●	●
(38) Characteristics of blood flow	●	○	○	○	○	●	○	●	●	●	◐	●	●
(39) Local phase quantization	●	○	○	○	○	●	○	●	●	●	◐	●	●
(40) Automatic adaptation to new spoof materials	●	○	○	○	○	●	○	●	●	●	◐	●	●
(41) Wavelet-based detection	●	○	○	○	○	●	○	●	●	●	◐	●	●
(42) Image power spectrum	●	○	○	○	○	●	○	●	●	●	◐	●	●
(43) Using thin-plate spline distortion model	●	○	○	○	○	●	○	●	●	●	◐	●	●
(44) Algorithmic-based counter measures	●	○	○	○	○	●	○	●	●	●	◐	●	●
(45) Texture and 3D structure analysis	●	○	○	○	○	●	○	●	●	●	◐	●	●
(46) Sparse low rank bilinear discriminative model	●	○	○	○	○	●	○	●	●	●	◐	●	●
(47) Locally uniform comparison image descriptor	●	○	○	○	○	●	○	●	●	●	◐	●	●
(48) Fourier spectra analysis	●	○	○	○	○	●	○	●	●	●	◐	●	●
(49) Optical flow and structure tensor analysis	●	○	○	○	○	●	○	●	●	●	◐	●	●
(50) Using local descriptors	●	○	○	○	○	●	○	●	●	●	◐	●	●
(51) Demodulation by complex-valued wavelets	●	○	○	○	○	●	○	●	●	●	◐	●	●
(52) Using quality related features	●	○	○	○	○	●	○	●	●	●	◐	●	●
(53) Morphology analysis of physiological signals	●	○	○	○	○	●	○	●	●	●	◐	●	●
(54) Correlation of physiological signals	◐	◐	◐	◐	◐	●	◐	●	●	●	◐	●	●
(55) Fusion of physiological signals	◐	◐	◐	◐	◐	●	◐	●	●	●	◐	●	●
(56) Eulerian video magnification	●	○	○	○	○	◐	○	◐	◐	◐	○	◐	●
(57) Alignment of audio and video	●	○	○	○	○	◐	○	◐	◐	◐	○	◐	●

● = **High** ◐ = **Medium** ○ = **Low**

potential solutions.

3.11.1 Ideas based on Existing Works

Modern techniques are mostly derived from challenge-response methods. For example, while the subject's trait is being captured, she will be asked to read out a random text displayed on the screen. In another scenario, in case of facial images or brain EEG signals, subject will be asked at random times to blink. Observing and detecting the impact of blink on the signal at that exact moment is a sign of liveness. The randomness of the moment of the stimuli occurrence and also its type, is the essential trick here to prevent or at least harden forged traits. The stimuli can even be more hidden; an ultra sound can be played which the human ear cannot hear, but effects the bio-electric signals. These kind of techniques are very effective in supervised settings but in an unsupervised settings the adversary might be able to have a real-time signal generation module which is capable of mimicking the effect of the stimuli and then adding it the previous forged signal. Especially considering the capabilities of computer vision and voice recognition/generation systems, such a system can help the adversary automate the process of sensing the upcoming stimulus and integrate their effects in the forged signals or generating the extra signal needed. A counterattack could be using CAPTCHA techniques to prevent the stimuli request being understandable for the machine but still comprehensible to human subject. For example, a simple idea would be to ask subject to read out the text inside a CAPTCHA or give voice commands to subject in a fashion only understandable to human.

Another idea would be to ask the subject to do an unfashionable/strange move (e.g. rising eyebrow, or shaking head), which its effect on the input is not known to adversary, so it would be infeasible to have a ready module to add its impact to the base crafted input. One might argue how the original system can use this strange stimulus for liveness detection, and one answer can be that in this scenario the stimulus itself is not important, but the knowledge of system of the exact time to expect some kind of extraordinary change in signal is the sign of liveness. In other words, deliberate noise addition can be a sign of liveness or even a possible feature for authentication. Furthermore, human errors and mistakes can be a way of distinguishing her from error-free machine, as an example, asking complex questions out of a subject field of study, e.g. chemistry question from a computer engineer, which she is supposed to not be able to answer it.

3.11.2 Psychological-based Interaction

The general idea here is to interact with the subject and analyze the response as a sign of liveness. These types of techniques try to distinguish between human and machine intelligence, and are different from methods that analyze voluntary/involuntary subject's reflections such as pupillary response to light or eye closing request. Some examples of these techniques is listed as below:

- **Emotional Reaction**: checking how the subject react to an emotional stimuli e.g. a scary movie or a humorous joke.
- **Human Errors**: usage of relative visual perception and optical illusions that lead to human error, but machine can solve it error-free. For example, Figure 3.4 shows an example puzzle question where machine and human intelligence will give different answers.
- **Rapid Human Learning**: Teach and ask at the same time to leverage human capability in learning quickly while machine training is time consuming. For example, human can read this unknown sentence, while machine may have

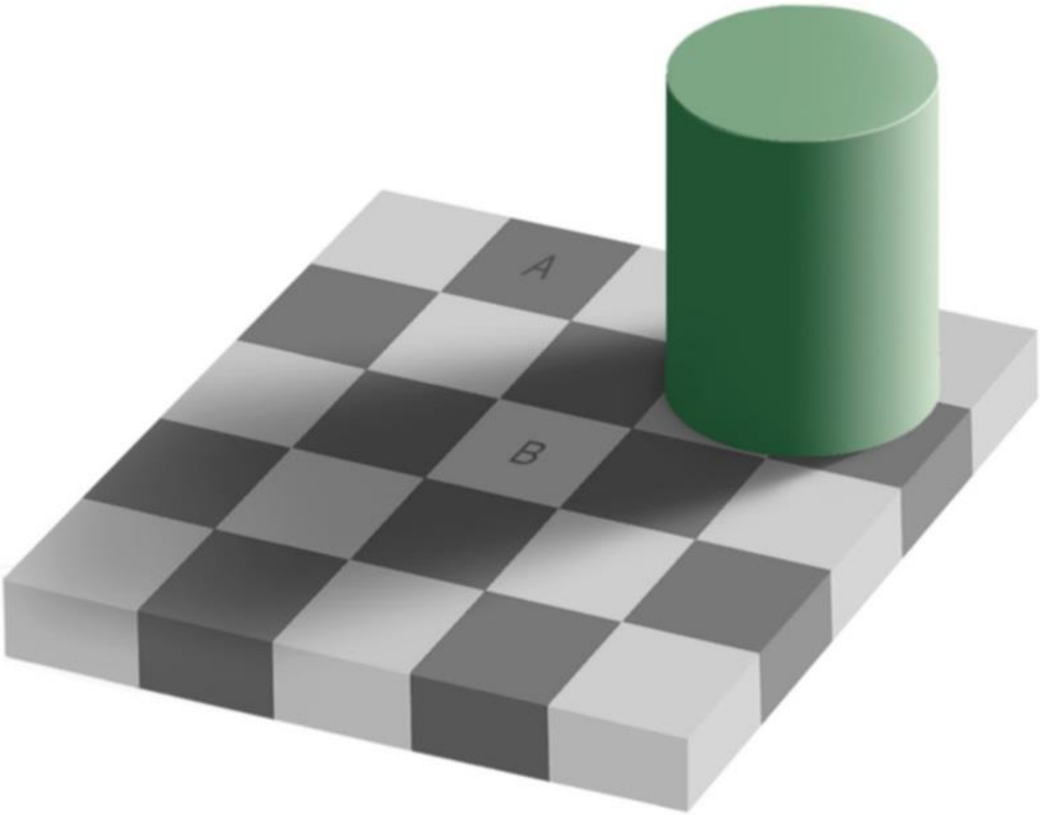


Figure 3.4: Which box is darker? A or B? To Human Eye, Box A Seems to Be Darker, However the Two Boxes Are Actually the Same Color and Machine Vision Systems Would Also Say So. An Example of Utilizing Human Weaknesses for Liveness Detection (net, 2019a)

problem: ”70 B3, oR No7 7o B3: 7H47 15 7H3 QU3571ON”.

- **Specialized Questions**: asking complex questions out of subject’s field of study e.g. medical question from a computer engineer. She is supposed to not be able to answer it.
- **System History**: questions about previous interactions with the system (e.g. what was the last setting you changed in the system options?) There can be

some analogy between this idea and the questions asked for recovering forgotten passwords on online services (e.g. email accounts), where subject has to answer to a set of questions defined by herself during preregistration. If the answers match with the previously given answers, the subject can set a new password for her account.

- **Attacking the attacker**: exploiting the weak-points of attacker. For example, CAPTCHA can be used in a reverse manner, where it is easy for machine, but hard for human to understand, as shown in Figure 3.5. In this case, during LD only machine language (the first line in Figure 3.5) will be displayed. If the subject enters the password, it is a machine.

- **Personality Tests**: Using psychological test which target user's subconscious or unconscious. For example, Hungarian psychiatrist, Leopold Szondi developed a test that asked which person in Figure 3.7, you do not prefer to meet at night? The respond would reveal some characteristic about the patient's personality. These kind of questions with answers rooted in our deep unconscious mind can be the exploited for liveness problem.

These are just some of the ideas for psychological challenge-response based techniques, and each one can be elaborated in various ways.

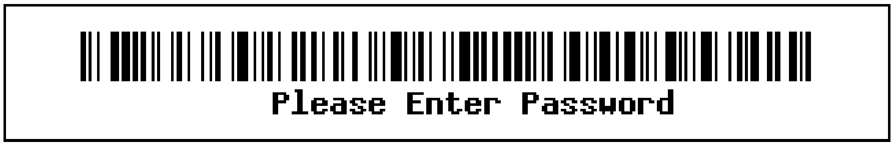


Figure 3.5: Machine vs. human language (image created using net (2019c)).

3.11.3 Biometrics Correlation

In multi-modal sensing, considering traits are independent from each other, correlation between the trait (either medical or mathematical correlation) can be exploited. In this case, if the sorted security bits for traits are $\{S_1, S_2, ..., S_n\}$, potentially system can reach security bit of $S_1 + S_2 + ... + S_n$ for the whole system, which is a significant increase. For example, the correlation between ECG and ABP Cai and Venkatasubramanian (2016) can be used to extract a more complex security feature which is harder for the adversary to regenerate and present it to the system as live data. In this strategy, the adversary should guess a biometric trait first, then guess a second biometric trait with respect to the first one. In the best case, the security strength will be the sum of the security of each biometric, but usually it is lower than the sum in practice Takahashi and Murakami (2014). Figure 3.6 shows how two correlated biometrics can decrease the chance of successful adversary guess by reducing the acceptance feature space. Different types of correlation can be found between two biometrics such as *temporal* and *morphological* correlations. In the former, the time and order of significant events in traits are studied, while in the former, shape and representation of the traits and their matching points are analyzed.

3.11.4 Biometrics Fusion

Multi-modal fusion where more than one trait is captured from subject is another possible solution for liveness problem. It may not solve the problem, but significantly can increase the robustness of the system against presentation attack. As seen in Figure 3.8, fusion can be applied in different system levels Ross and Jain (2003): 1) analog data fusion, 2) digital data fusion Cai and Venkatasubramanian (2016), 3) feature fusion Nagar *et al.* (2012), 4) matching score fusion Takahashi and Murakami

(2014), and 5) decision results fusion. For example, in the last approach, each trait should pass the liveness test, and then the result of these tests are integrated to reach the final decision. In this case, if the sorted security bits for traits in ascending order are $\{S_1, S_2, ..., S_n\}$, then the max achievable security bit for the whole system would be $S_n + 1$.

3.11.5 Continuous Monitoring

BAS records subject's trait in an ongoing manner while they are interacting with system to enable progressive authentication procedures. The stream of data can be utilized to extract the degree of confidence in subject's identity which would determine their access level especially in case of security critical tasks such financial payments or changing other users' access level. Continuous monitoring would also allow for building a model for subject's habits and behaviors through time. Later on this comprehensive understanding of subject's characteristics can help with designing

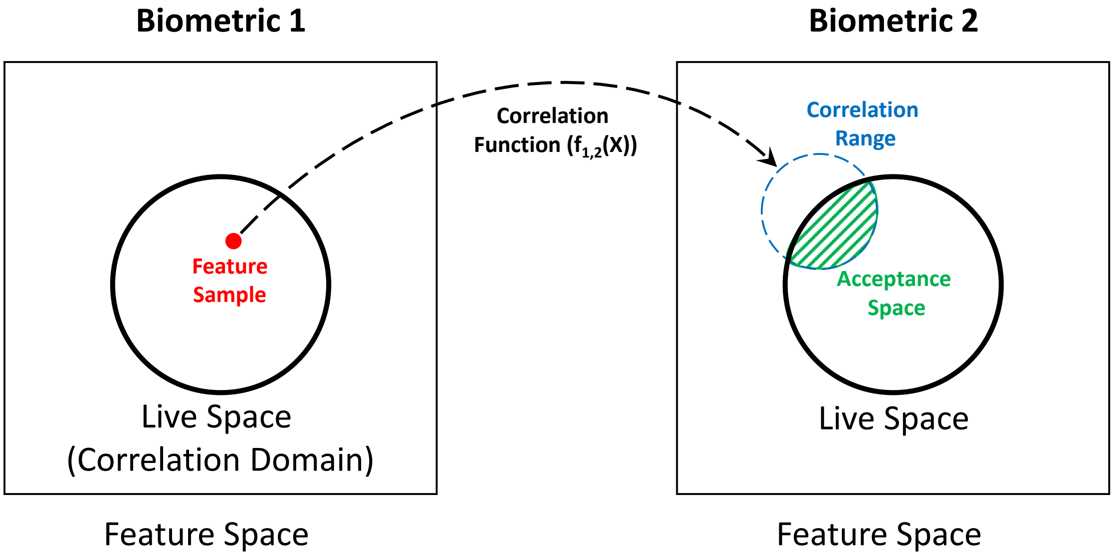


Figure 3.6: Reduction in attacker choice using correlation.

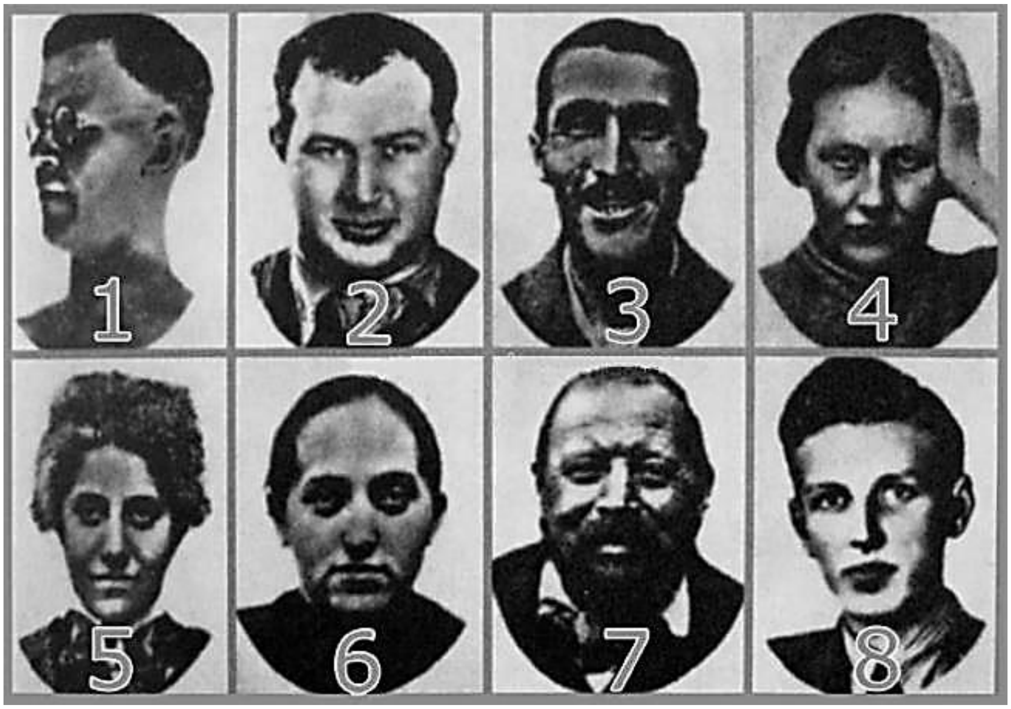


Figure 3.7: Which person scares you the most? net (2019g)

personalized challenge-response tests to distinguish her from human/machine adversary. For instance, system's interpretation of the collected data might suggest user prefers colored themes over dark/white ones for system user interface. So a correct respond from the subject to the question about her preferred themes during authentication process can be used for ensuring liveness. However, there can be some privacy concerns with continuous monitoring as the recorded data (if compromised) can be used by adversaries to learn and understand (some other) subject's behaviors. For example, Electromyography (EMG) which records the electrical activity of muscles which is used in motion detection can be exploited to detect subjects key strokes on keyboard (Zhang *et al.*, 2017). Hence, special attention should be paid to privacy implications of recorded data in one-time or continuous settings.

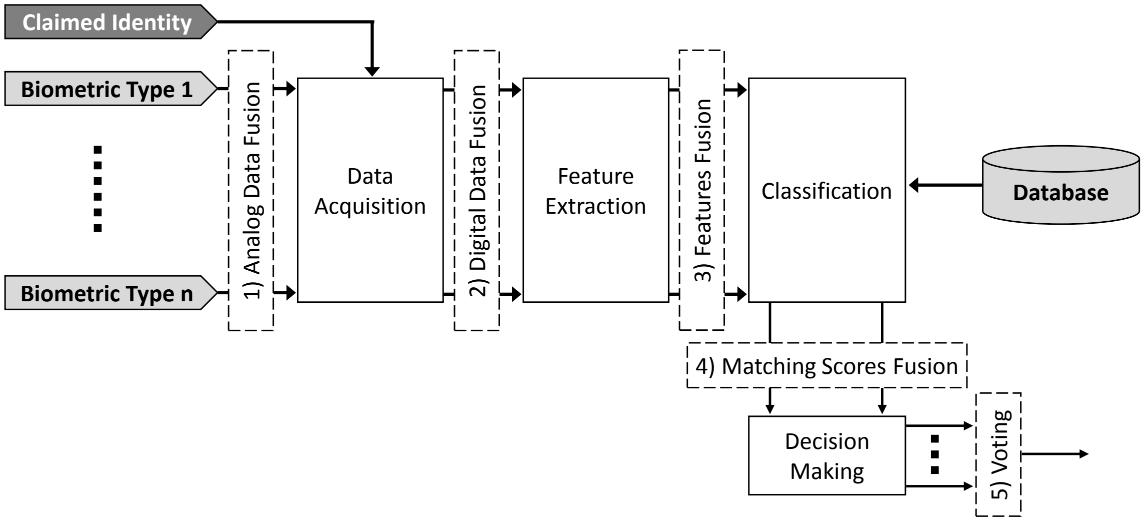


Figure 3.8: Different levels of biometric fusion.

3.12 Discussion and Future Approach

As discussed in section 1.0.1, liveness test can be studied through the fundamental concept behind Turing test; discerning between a human and a machine. Emergence of complex presentation attacks against human traits, in addition to advances in Artificial Intelligence (AI) domain which allows machines to generate high quality contents (image, voice, video, text, art) has implications for liveness and Turing test. These meticulously forged data can spoof both liveness test, where machines attempt to discern between a human and a machine, and Turing test, where humans should discern between a human and a machine.

If liveness test fails, changing the interrogator from machine to human can generally resolve the issue as in case of using human instead of machines to detect gummy fingers or face masks. Basically, human is acting as the ground truth in such cases. Although there can be scenarios such as with artificially generated brain (EEG) signals which a human expert does not have a significant advantage over machine or even have a disadvantage.

If Turing test fails replacing the human interrogator with a machine would not necessarily be more effective as with AI-generated text. In cases such as of detecting some types of deepfake videos, machines might have advantage over humans as they analyze the digital representation of frames and not the visual representation as by human. A primary challenge is related to the point that in computer systems, only human experts can deterministically decides on ground truth. In situation where human expert fails to do so, the decision made by the machine is inevitably non-deterministic as there is no entity left to verify it. This is unlike the case with failure in liveness test which human can intervene as the ground truth.

In other words, machines should not be capable of passing the liveness test and more importantly the Turing test. There is a direct relation between the two tests and their performance depends on each other. Hence, research is necessary for both tests in order to design effective approaches and methods for discerning between human and machine.

Currently, there are programs such as ELIZA and PARRY which will pass the original chat-based Turing test. Therefore, updated Turing tests should be designed in a manner which the human interrogator can easily distinguish between a human and a machine. For this objective, the new Turing tests should be checking for tasks which current and foreseeable machines cannot complete but human is capable of accomplishing these tasks. CAPTCHAs (Completely Automated Public Turing test to tell Computers and Humans Apart) can be a appropriate starting point for these kind of tests. Moreover, machines capabilities for understanding and generating contents in relation to vision (image/video), sound (voice/audio) and language (comprehending/responding) is mostly based on Machine Learning (ML) models which are not completely robust. Some studies has shown that for object detection in images, slight changes to pixels can result in misclassification. These changes are not recognizable by human eye, but will interfere with the machine's operation. Therefore, this type of images which fools the machines but not humans, would be a possible candidate for Turing tests and even liveness test. Although such weakness is beneficial for this goal, one should not to forget that misclassification of a stop sign by an autonomous vehicle poses serious dangers and can cause accidents. Basically, any shortcomings in a machine's performance which researchers try to resolve in other domains, can potentially be utilized in Turing/liveness tests to help distinguish between human and machine. At the same time, advances in other domain while it will ease human life from some aspect, it will also facilitate creating more advanced adversarial machines.

Beside machine's shortcomings, those of human can also be leveraged. In this approach, a test will be used that machine will produce the correct response to it, but the human brain will output an incorrect answer. Basically, it is not necessarily needed that either of human or machine produce correct answers, just that they produce different answers would be sufficient. An example would be using optical illusion; human brain believes something which is not true, however the machine will process the image and just reach the actual facts. Therefore, the human and machine will generate different answers, which it would be an indicator for discerning between them. In other words, machines do not make mistakes while humans do, and this human imperfection can be used for liveness detection.

After all this discussions on various approaches for liveness detection, one might ask is there any final solution to liveness detection? The theoretical answer is yes, while the practical answer is no (at least by the current state-of-the-art). From the field of computational complexity theory, it is known that while NP problems can not be solved in polynomial time, a candidate answer can be verified in polynomial time. Therefore, theoretically a final solution can be achieved if there exists a human trait (or behavior) which replicating it would be a NP problem for the adversary. This trait would be optimal and unforgeable since the adversarial machine is incapable of artificially generating it in polynomial time, while the liveness test can verify if its an original trait in polynomial time. Basically, any problem related to human body, mind or intelligence level that has NP problem characteristics on average case would be the ideal candidate for liveness detection.

As of now, no human trait is known to be NP problem so a prefect solution can not be achieved and instead focus should be on the best practical solution. To design an optimal authentication system which should be usable, real-time, private and secure, one should not solely rely on one class of techniques (password, security token, human trait) but instead on a mixture of them. The traditional factors in authentication can be summarized in the well-know expression of “Something You Know, Something You Have, or Something You Are”, but for enhanced robustness the approach should be “Something you know and you have and you are”. Beside fusing these three factors, new type of factors should be researched, as some suggest utilizing user location as another factor.

3.13 Conclusion

Alongside with the widespread integration of biometric authentication systems in different domains of human life, numerous presentation attacks against such systems has emerged leading to security and privacy of users to be compromised. Liveness detection methods have not matured with the accelerating pace of using human traits for authentication task, and the ever increasing emergence of novel and complex attacks targeting biometric system. In this chapter, I presented a systematic analysis of the state-of-the-art in liveness detection domain to help bridge the gap between attack and defense methods. I provided a detailed taxonomy of liveness detection for human traits (e.g. fingerprint, face, voice, iris, vein, heart signal) to facilitate integration of successful methods into designing more complex methods and allow for effective methods to be adapted in case of other traits. Moreover, by investigating the literature, I observed an serious flaw in design methodology of biometric systems. Currently, most biometric authentication systems are designed in isolation from liveness detection modules and then simply the two are being merged. In a worse case, biometric system has been deployed and later on a liveness module is being deployed on the side to remedy its vulnerability against presentation attacks. In the worst case, biometric system is operating without any liveness module. A design shift is required to enhance robustness against presentation attack, and that would be to design the authentication and liveness detection modules in relation to each other and not separately.

Previously, liveness methods were only evaluated based on their performance, however I proposed two more dimensions to be considered: usability and cost. Afterwards, I evaluated the state-of-art in liveness detection methods with respect to this three dimensions. In case of performance, I suggested several threat models and methods were evaluated against all of them. To the best of my knowledge, this is the first work to propose an evaluation criteria and metric to allow for more precise comparison of different methods. Of course, as with any first attempt there is plenty of room for improvement and designing more comprehensive evaluation criteria and metrics.

Finally, I argued that liveness detection methods should not only focus on biological characteristics of human traits, but also investigate cognitive and intellectual aspects of humans for ensuring liveness of the entity interacting with biometric system. Moreover, I proposed a series of new approaches and solutions that target such features of humans in contrast to machines. Basically, I attempted to study the liveness detection problem from a new angle and discuss it in the broader framework of how a machine can distinguish between a human and a machine (i.e. liveness test). This is important since with advances in artificial intelligence and robotics, human society is reaching the era of autonomous machines and researchers should make sure such machines only accept command from humans and not other adversarial machines.

Chapter 4

BRAIN-BASED AUTHENTICATION SYSTEMS

In this chapter, I will discuss my research in relation to brain-based security systems. I started by proposing E-BIAS (EEG-based Identification and Authentication System), which was one of the early works to use commercial brain senors instead of medical ones (Sohankar *et al.*, 2015a). Afterwards, a novel presentation attack using artificially generated signals was proposed which successfully bypassed 30 different authentication configurations; five feature extraction methods and six classification algorithm (Sadeghi *et al.*, 2017). The success of the proposed attack challenged the *intrinsic liveness* property which is presumed by literature for brain signals (Zhao *et al.*, 2019; Kong *et al.*, 2018; Maiorana and Campisi, 2017; Garau *et al.*, 2016; Thomas and Vinod, 2016; Sundararajan *et al.*, 2015; Fraschini *et al.*, 2014; Nakanishi *et al.*, 2009). In another work with application in Brain Computer Interaction (BCI) systems, visual stimuli was presented to subjects which was expected to increased stress levels. Impact of stimuli on brain signal as the primary trait and on heart signal as the secondary trait was detected (Sadeghi *et al.*, 2016a). This technique to add context (e.g. stimuli) and detect its expected impact on brain signal and the correlated heart signal, can be utilized for ensuring live and timely properties of an input EEG signal in an authentication system.

4.1 E-bias: EEG-based Authentication and Identification System

Brain sensing and associated cognitive applications are fast becoming pervasive in nature due to the advent of wireless low cost easy-to-wear brain sensors that connect to mobile phones (Campbell and Choudhury (2012); Oskooyee *et al.* (2014)). This enables seamless access to a person's brainwaves which contains information that is unique to a person, nearly impossible to impersonate without invading personal space, and chaotic over time. This is markedly different from biometrics such as fingerprints, voice, and face, which can be captured without the subject's knowledge or purposefully altered (Hu *et al.* (2011)). Seamless availability of EEG data opens up potential usage in securing personal information in scenarios where a password may not be entered, spoken out, or remembered. For example, the notion of "hands-free" security can be imagined, when the person is driving or pre-occupied with other tasks and cannot focus on targeted security related tasks (Banerjee *et al.* (2013)). EEG-based security systems satisfy the following requirements that favor the mentioned purpose: a) universality, we always have our brain and thoughts with ourselves and hence enables pervasive security, b) uniqueness, brain signals are unique and differ from person to person potentially enabling high authentication or identification accuracy, c) permanency, some brainwave features show stable underlying behavior through time, which can be classified using machine learning techniques, but are difficult to regenerate without prior access to brain data (Lee *et al.* (2013)), d) collectability, they can be captured by wearable sensors, and e) robustness, it's hard to hack a system through replication of brain data (Almehmadi and El-Khatib (2013); Zúquete *et al.* (2011)). We propose a seamless pervasive EEG-based security system using commercially available brain sensors intended to provide authentication and identification in single user smartphones and small scale multi-user computing systems. The E-BIAS is distinguished by relatively lower training time than existing techniques and with simple mental task for the user. In our system, we use single channel commercially available EEG headset, Fast Fourier Transform (FFT) for feature extraction, naïve Bayes classifier (NBC) for classification, two minutes training time, and 10 test subjects. The authentication accuracy ranges between 81-95% depending on the length of testing samples (5-60 seconds of EEG samples), and the maximum identification accuracy reaches 80% with 50 seconds EEG test samples.

4.1.1 System Requirements, System and attack Model

Previous works in pervasive interactive applications (Ferreira *et al.* (2013)), brain monitoring (Sharieh *et al.* (2008); Zao and et. al (2014)), and biometric systems (Almehmadi and El-Khatib (2013); Khalifa *et al.* (2012)) address a number of system requirements that can be used to evaluate a pervasive EEG-based security system:

1) Accuracy: The chaotic nature of brain signals and impact of varying emotional states (anxiety, stress, anger, etc.) or drugs on the EEG signals can affect the success rate of security system over time. The present work uses baseline EEG signals, when a person is in rest state which is shown by recent works (Almehmadi and El-Khatib (2013); Lee *et al.* (2013)) to remain stable over long periods of time. In this state, it is difficult to extract unique features of a person especially using commercial sensors. There are two possible solutions: 1) applying complex preprocessing, feature extraction, and classification methods, which might increase power consumption, and response time, and 2) collecting large training data set that can dramatically increase response time. In our research, we employ the first method. We use Fast Fourier Transform (FFT) for feature extraction and machine learning method called Naïve Bayes Classifier (NBC) that leads to high accuracy performance.

2) Timeliness: Processing EEG data using complex machine learning techniques is a time-consuming procedure especially on mobile platforms. Based on the experimental results, our system runtime on mobile phone is approximately 100 times slower than running on a desktop system. Therefore to avoid unbearable latencies as well as fulfilling real-time requirements, we use a "fog sever" based system architecture, that enables the smartphone to offload complex data processing for faster execution time (Pore *et al.* (2015); Zao and et. al (2014)).

3) Energy Efficiency: Mobile platforms provide limited amount of resources such as energy, bandwidth, and storage capacity. The complex computation required for our system, drains the smartphone battery. Here again, using fog server (i.e., laptop) in the system architecture saves smartphone energy, and help the system handle computational and storage requirements for the application.

4) Usability: It is the ease of use of system while not deteriorating its performance.Tedious training procedures reduces system usability. Various tasks with roots in psychology and neuroscience have been designed and exploited in different works. For instance, resting/relaxing, imagining moving body parts, auditory/visual stimulation (e.g. tones, songs, colors, or images), performing mathematical operations in mind, thinking about a specific concept (Chuang *et al.* (2013); Marcel and Millán (2007)), or even without doing any tasks (Hu *et al.* (2011)). In our research, to keep the scenario simple and acceptable by the user, we use brain signals while the user is in physically rest state. We also use a light-weight commercially available wireless EEG sensor with a single dry electrode that records signals from forehead. At last, mobile platform is used to implement sensor interface in pervasive contexts.

5) Robustness: The system should maintain required levels of security under various attacks. We evaluate our system against three types of attacks: a) impersonation, where a person attempts to imitate another person's EEG signals, b) database hacking, where the unique brain signature of a person is stolen, and c) communication snooping, where the brain features transmitted from a user over network are stolen. We use ten subjects to test the system that is comparable with recent research.

EEG Signals

EEG signals are electrical flows through neurons caused by brain activities which produce potential differences in order of microvolts (5-100 μV). EEG signals are captured by placing EEG electrodes on the surface of scalp (Tan (2006); Wolpaw *et al.* (2002)).EEG signals are usually decomposed in several frequency bands. Each band contains signals associated with particular brain activity (Chuang *et al.* (2013)): 0.5-3.5 Hz (δ, sleep state), 4-7 Hz (θ, drowsy state), 8-13 Hz (α, relaxation or rest state), 14-30 Hz (β, active concentration and alertness state), 30-100 Hz (γ, perception). In our experiments, we consider only the rest state which is marked by large variations in α wave amplitudes and it is achieved by requiring each subject to sit on a chair and relax in a distraction-free room.

EEG waves are considered to be deterministically chaotic signals (Niedermeyer and da Silva (2005)). This means that their amplitude and duration are highly random but their unpredictability can be mimicked by non-linear dynamic learning systems such as neural networks. The significance of such chaotic nature for security system is that given a sample data set S_i and a non-linear dynamic learning system M it is extremely difficult to derive another sample data set S_i' that is accepted by the system M (Barreno *et al.* (2006)).

System Model

Figure 4.1 shows system model of a mobile phone security system using the proposed EEG-based solution. In this model, a mobile phone collects brain sensor data, and sends it to a fog server for extraction of EEG features and classification. The fog server uses a cloud database for the purpose of classification. The cloud database stores EEG signatures/features from each of potential users that can log into the

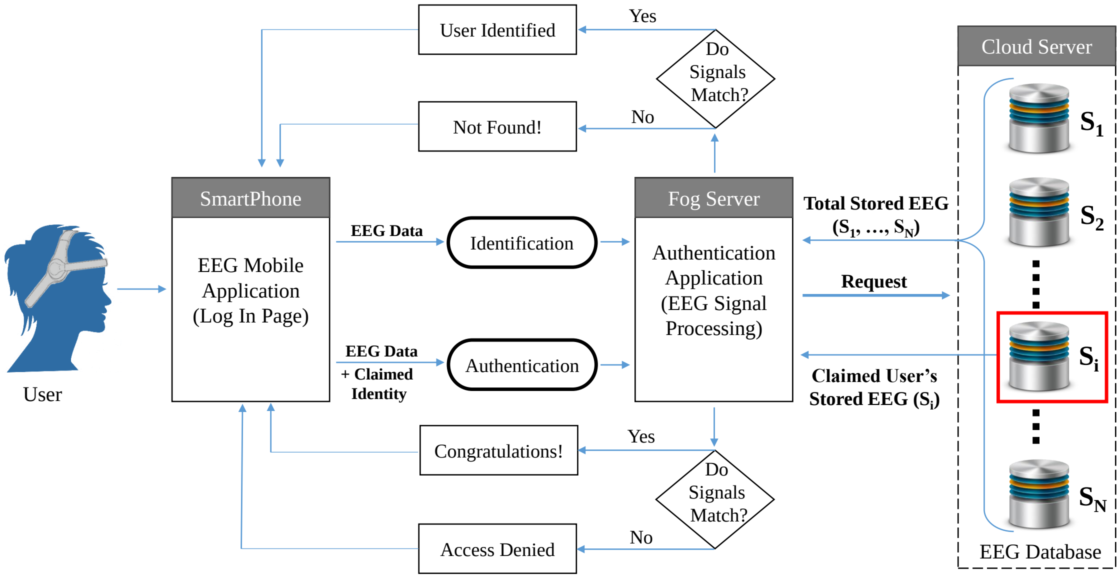


Figure 4.1: Model of EEG Driven Security System.

system. The application can run in either the identification or the authentication modes which are defined below.

Authentication problem: Considering a pair of signal and identity, system should specify whether the signal matches the stored signature of the identity.

Identification problem: Considering an input signal, system should specify whether the signal matches any of the stored signatures in the system.

Identification is a more complex problem than authentication due to two main reasons: a) in identification we have to search through features of multiple subjects while in authentication, our search space is restricted to only features of a given subject, and b) the identification problem has to handle cases when input features may get classified as signatures of more than one subject.

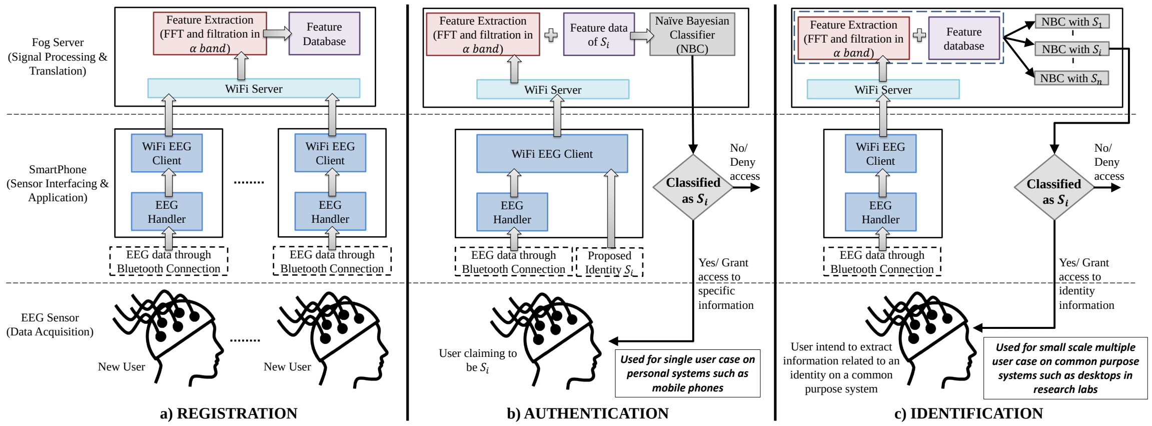

Fog Server
(Signal Processing & Translation)
SmartPhone
(Sensor Interfacing & Application)
EEG Sensor
(Data Acquisition)
Feature Extraction (FFT and filtration in α band)
Feature Database
WiFi Server
WiFi EEG Client
EEG Handler
EEG data through Bluetooth Connection
....
New User
New User
a) REGISTRATION
Feature Extraction (FFT and filtration in α band)
Feature data of S_i
Naïve Bayesian Classifier (NBC)
WiFi Server
WiFi EEG Client
EEG Handler
EEG data through Bluetooth Connection
Proposed Identity S_i
Classified as S_i
No/ Deny access
Yes/ Grant access to specific information
User claiming to be S_i
Used for single user case on personal systems such as mobile phones
b) AUTHENTICATION
Feature Extraction (FFT and filtration in α band)
Feature database
NBC with S_1
NBC with S_i
NBC with S_n
WiFi Server
WiFi EEG Client
EEG Handler
EEG data through Bluetooth Connection
Classified as S_i
No/ Deny access
Yes/ Grant access to identity information
User intend to extract information related to an identity on a common purpose system
Used for small scale multiple user case on common purpose systems such as desktops in research labs
c) IDENTIFICATION

Figure 4.2: System Architecture for Different Usage Scenarios.

Usage Scenarios

We envision the usage of the proposed EEG-based security technique in two scenarios: a) authentication of an individual to a single user personal mobile device (smartphone), and b) identification of an individual as a registered user for a small scale multi-user computing system such as common purpose desktops in a research facility as seen in Figure 4.2. We will refer to the system to which the user wants to get authenticated to as the target. The first step in both scenarios, is registration procedure.

Registration: In the registration process (Figure 4.2(a)), a user (S_i) is required to wear the Neurosky headset and the target collects a 2 min sample of EEG data. During these two minutes, the user is required to be in *rest* state while the user is not doing any specific mental task. The 2-minute sample is then passed to a fog server, with higher computational capacity, for extracting relevant features and storing them in a database. According to our experiments, when the target is a smartphone, it typically needs an external desktop system as the fog server for fast feature extraction. The registration process is the same for a multi-user scenario and the only difference is that the fog server, stores a database of features corresponding to different users.

Authentication: Authentication process (Figure 4.2(b)) is intended towards a target, which only has a single user. In this scenario, the returning user who wants to gain access to the target registered to S_i, wears the brain sensor such that the target can collect a 1-minute sample of the brain signal and send it to the fog server. The fog server, now uses the feature set of S_i and compares the collected data from the returning user by applying machine learning techniques such as NBC. If the fog server can classify the returning user as S_i, then it grants access.

Identification: The identification process (Figure 4.2(c)) is intended towards a target, which has multiple (∼10) registered users. In this scenario, the returning user again wears the headset and the target collects a 1-minute sample. It then iterates over all possible registered users and uses machine learning techniques to classify. If the target achieves a unique classification, it grants the returning user access to the identified user account. In case of multiple or inconclusive classification the target denies access.

Trust and Attack Model

In a security system, no communication link (i.e. Bluetooth and WiFi networks) or computational units (i.e. mobile devices and fog servers) is completely secure. We consider three types of attack against our system and later on propose a solution to them. In these attacks, we consider the cases where the attacker tries to fake brain signals, hack computational units and monitor communication links.

Attack Scenarios

Impersonation: As shown in Figure 4.3(a), impersonation occurs when the attacker wears the EEG headset and tries to mimic a user's brain signal and fool the system into granting access to her.

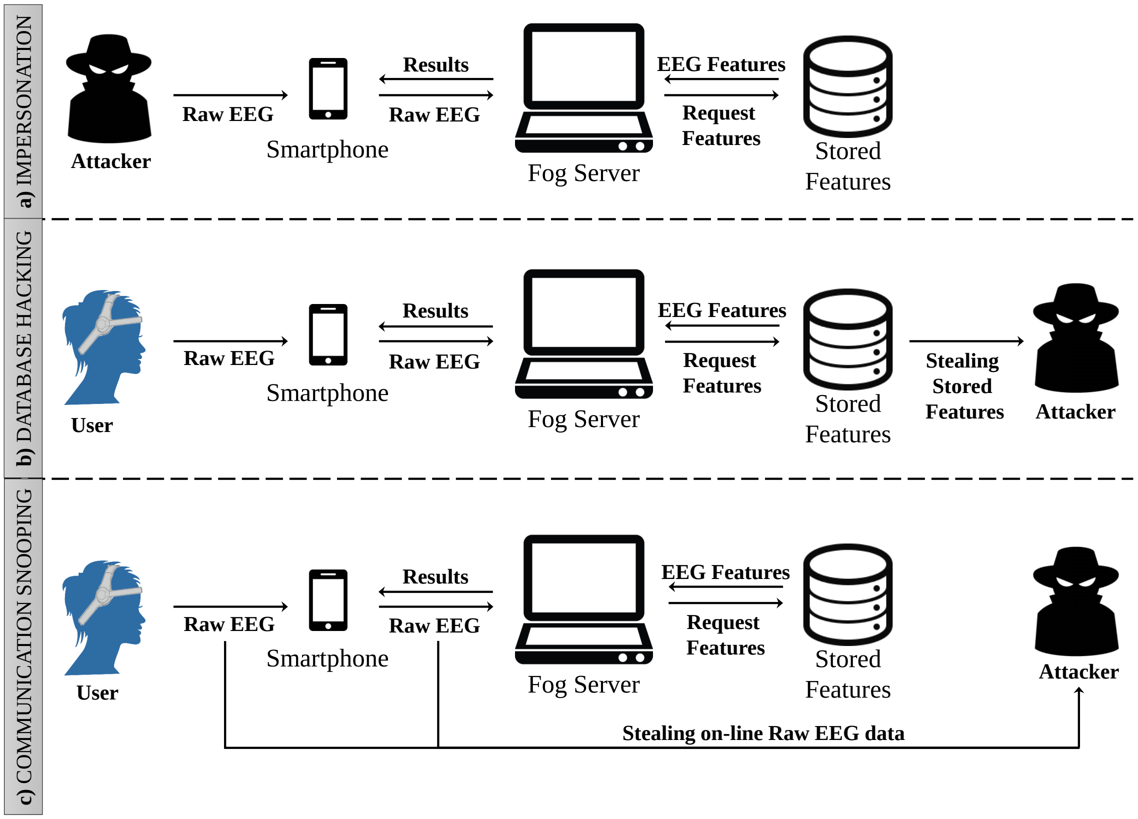


Figure 4.3: Three Type of Attacks Against System.

Database Hacking: In this type of attack, the attacker hacks the database of stored features which are the users signatures as in Figure 4.3(b). Then, the attacker can provide these features to the system and gain access.

Communication Snooping: The attacker monitors the communication links (e.g. between the EEG headset to smartphone or smartphone to fog server) and steals the current signal and features that user is providing to system, as seen in Figure 4.3(c). Later on the attacker can use these features to gain access.

4.1.2 System Evaluation

In this part, system components are evaluated with respect to accuracy and robustness.

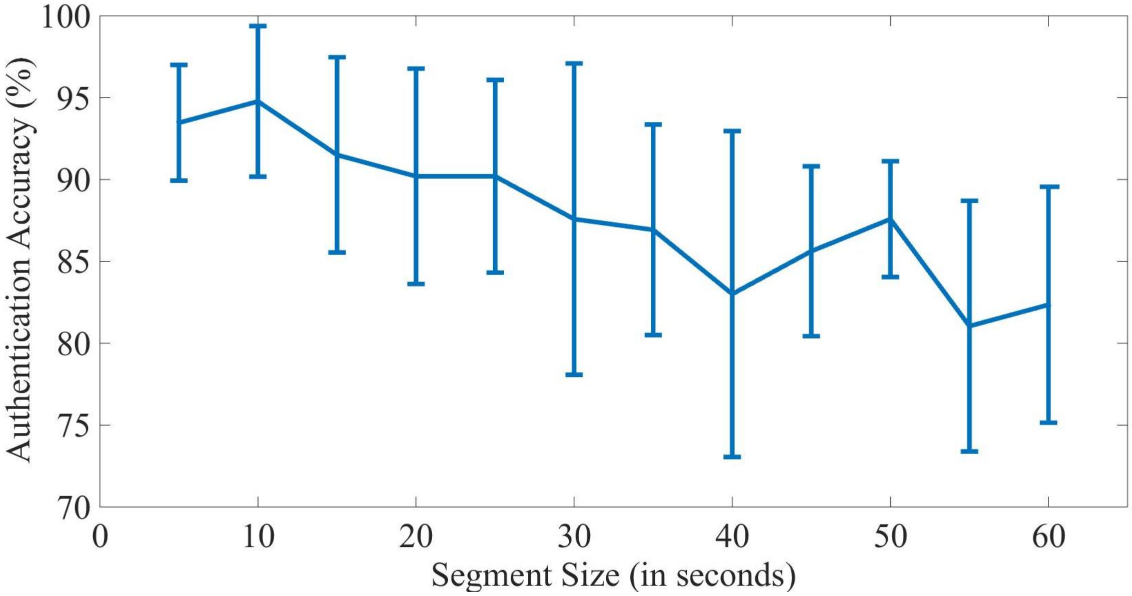


Figure 4.4: Average Authentication Rate Vs. Training Size (i.e. Segment Size in Seconds) over All Subjects.

Scenario description

In the experiment, we used a simple task for the sake of usability. In some works, they use scenarios which needed extensive effort from the subjects, but we just asked the subjects to be in rest state (stay calm and relax) for 2 minutes. Related works choose EEG recoding times range from 24 seconds up to 16 minutes. In our study, 2 minutes was long enough to reach desired performance and was short enough to avoid user frustration. The subjects were in silent room siting on a chair and without performing any specific mental task. We used NeuroSky mindwave sensor for capturing EEG signals from the subjects. For each subject, one session was captured. The subjects were graduate students of Arizona State University between the ages of 20 to 30 (both male and female). We had 10 subjects in our experiment which is comparable to other works. Each session data is divided into segments. We tested our system on different segment sizes from 5 to 60 seconds with 5 seconds interval.

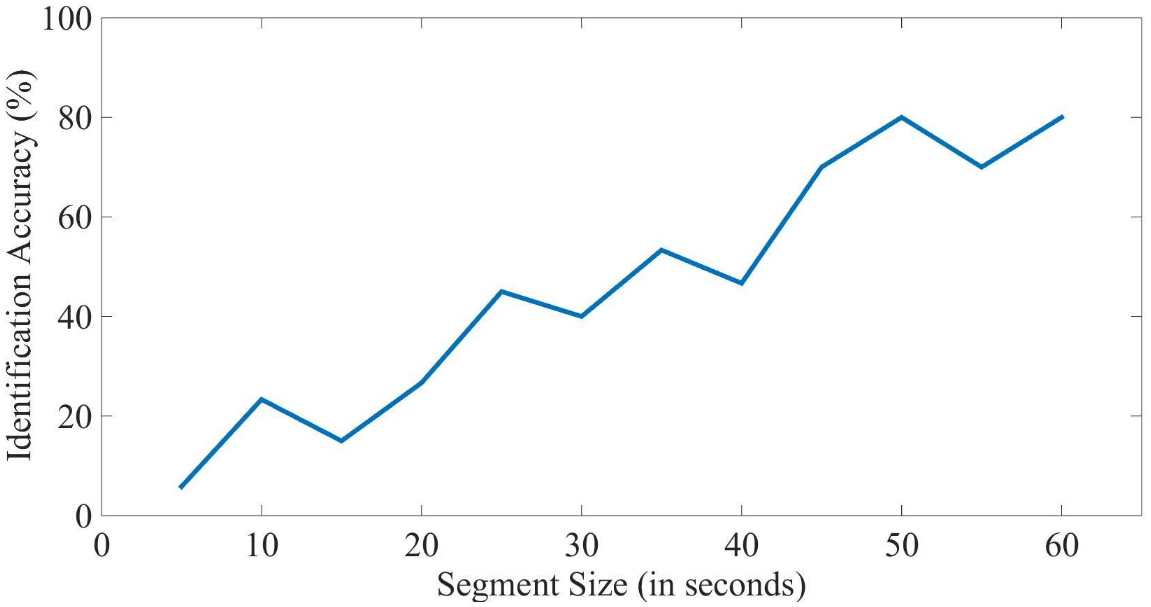


Figure 4.5: Identification Rate Vs. Training Size (i.e. Segment Size in Seconds) over All Subjects.

Authentication

To test the authentication mode of the system, we divide each subjects signal into segments. We choose one of these segments as the subject signature for our system database and use the rest of segments for testing the system. For each subject, we calculate TA, FR, TR, and FA. We check each segment of a subject with all the segments of the same subject to calculate TA and FR. Afterward, we check each segment of a subject with all segments of all the other subjects for calculating TR and FA. Finally, the accuracy is calculated for different segment size varying from 5 seconds to 60 seconds with intervals of 5 seconds. In Figure 4.4, the average and standard deviation of the accuracy over all the subjects for different segment size is shown. We can see that the accuracy ranges from 81% at 55-second segment to 95% at 10-second segment.

Table 4.1: Authentication Results.

Reference	**Classifier**	**Channel(s)**	**Subject(s)**	**HTER**	**Accuracy**	**Scenario Duration**	**Device**	**Usability Index**
Riera *et al.* (2009)	FDA	2	40	10.9%	-	9-12 minutes	Medical (ENOBIO)	5.4
Ishikawa *et al.* (????)	Cosine Similarity	1	10	-	85-90%	12 minutes	Medical (BioSemi)	5.6
Nakanishi *et al.* (2009)	Spectral Distribution Analysis	1	23	11%	79%	30 minutes	Commercial	5.7
Riera *et al.* (2008)	FDA*	2	51	1.7%	-	8-16 minutes	Medical (ENOBIO)	5.8
Abdullah *et al.* (2010)	NNs	1	10	-	70-87%	5 minutes	Medical (g.tec)	6.1
Lee *et al.* (2013)	LDA*	1	4	-	87-100%	5 minutes	Medical	6.3
Hema *et al.* (2008)	NNs	3	6	-	95%	100 seconds	Medical (Bio Amps)	6.5
Ashby *et al.* (2011)	SVM*+ voting	14	5	3%	97-100%	150 seconds	Commercial (Emotiv)	6.9
Chuang *et al.* (2013)	Cosine Similarity	1	15	14%	85%	12 minutes	Commercial (Neurosky)	7.3
Present work	NBC	1	10	2-9%	81-95%	2 minutes	Commercial (Neurosky)	8.5

* SVM: Support Vector Machines, LDA: Linear Discriminant Analysis, and FDA: Fisher Discriminant Analysis.

Table 4.2: Identification Results.

Reference	**Classifier**	**Channel(s)**	**Subject(s)**	**Accuracy**	**Scenario Duration**	**Device**	**Usability Index**
Chuang *et al.* (2013)	Cosine Similarity	1	15	22%	12 minutes	Commercial (Neurosky)	3.9
Hema and Osman (2010)	NNs	3	15	60%	200 seconds	Medical	4.3
Poulos *et al.* (2002)	LVQ NNs*	1	4(+75 intruders)	76-88%	3 minutes	Medical	5.0
Hu *et al.* (2011)	NBC	1	11	66-100%	4 minutes	Medical (NeXus-4/Mind Media)	5.1
Mohammadi *et al.* (2006)	NNs	1	10	80-97%	24 seconds	Medical	6.1
Paranjape *et al.* (2001)	DFA*	1	40	79-85%	68 seconds	Medical	7.2
Present work	NBC	1	10	80%	2 minutes	Commercial (Neurosky)	7.5

* LVQ NNs: Learning Vector Quantization Neural Networks, and DFA: Discriminant Function Analysis.

Identification

Similar to the authentication mode, we divide the signals into segments and choose one of the segments as the subject signature for the system database. Then again we test all the segments from all subjects to calculate accuracy. We calculated the accuracy for different segment size varying from 5 seconds to 60 seconds with intervals of 5 seconds. In Figure 4.5, the average accuracy over all subjects for different segment size can be seen. The accuracy starts at 5% at 5 second segment size and increases with the segment size and reaches to 80% accuracy at 60-second segment size.

According to accuracy results shown in Figures 4.4 and 4.5, although the segment size increases, the authentication accuracy does not change a lot. But, increasing segment size leads to increase in identification accuracy, significantly. The graphs show that with segment size equal to 60s, authentication accuracy is 81% while identification reaches the highest accuracy point. Authentication is a special case of identification where the feature set just contains one sample. In this sense, identification accuracy cannot exceed authentication accuracy. So, at 60s segment size, both accuracies are around 80% where the system performance is at the highest level and remains stable from then.

Usability Analysis

There are several parameters involved in usability analysis of EEG-based security systems such as type of sensors, number of channels, number of subjects, experiment duration, HTER, and accuracy. To compare the usability of our system with related works, we define a metric called **usability index**. The index is determined by assigning weights to evaluation parameters. Higher weights are assigned to parameters with more important roles in usability of the system. Tables 4.1 and 4.2 list test setup parameters, results, and usability indexes for authentication and identification, respectively. All the experiments use rest task for EEG recording. In our study, we set the parameter weights (w_i) for usability index of authentication process as follows: "scenario duration": 3.0, "accuracy": 2.0, "number of channels": 2.0, "device type": 2.0, "number of subjects": 1.0, and "HTER": 0.5. In identification, number of subjects play an important role in the system performance, so higher wight is assigned to it, "scenario duration": 3.0, "accuracy": 2.0, "number of subjects": 2.0, "device type": 2.0, and "number of channels": 1.0. For usability analysis, number of subjects and accuracy are directly proportional to usability index, and we define their corresponding factors as seen in Equation 4.1:

$$f_i = \frac{x_i}{max(X)} \tag{4.1}$$

where X is set of values for a specific parameter and x_i is the parameter value for study i ($x_i \in X$). On the other hand, remaining parameters (i.e, number of channels, HTER, and scenario duration) and usability index are inversely proportional, as defined in Equation 4.2:

$$f_i = 1 - \frac{x_i}{max(X)} \tag{4.2}$$

In addition, the factor value for medical and commercial device types are set to 0.0 and 1.0, respectively. To calculate usability index, each weight is multiplied to its corresponding factor value, summed with other factor products, and finally divided by sum of the weights as seen in Equation 4.3. Higher usability indexes indicate higher system performance. As seen in the last column of Tables 4.1 and 4.2, the present work has higher usability index compared to previous works (for more readable representation, usability indexes are multiplied by ten).

$$Usability\ Index = \frac{\sum f_i w_i}{\sum w_i} \tag{4.3}$$

Security Analysis and Robustness

In this section, we evaluate our system against the three attack scenarios.

Impersonation: Robustness against impersonation attacks depends on FAR. We performed a study where 10 authentication systems were setup each for a given subject. For a particular authentication system, we attempted authentication using the EEG data from the other 9 subjects. Table 4.3 shows the number of times a wrong person was authenticated to a system out of 24 trials. The rows in Table 4.3 denote authentication systems for subjects 1 to 10, while the columns indicate authentication attempts from subjects 1 - 10. We derive two important conclusions from this experiment: a) the FAR is pretty low for most of the individuals, and b) the false accept events are independent. This means that false accept is rare and it is unlikely that a person gets consecutive wrong accesses (in most of the cases there were only 1 false accept while there is no evidence of consecutive false accepts). Hence, impersonation attacks can be countered by requiring multiple consecutive successful attempts. For example, if we require that the user has to have three consecutive successful authentication to gain data access then this reduces the FAR from 0.05 to $(0.05)^3 = 0.000125$, however, latency increases only linearly.

Table 4.3: Number of False Accept Events Out of 24 Trials.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
S_1	-	0	**1**	0	0	**1**	0	0	0	0
S_2	0	-	0	0	0	**1**	0	0	0	0
S_3	0	0	-	0	0	**1**	0	0	**1**	0
S_4	0	0	0	-	0	0	0	**1**	0	0
S_5	0	0	0	0	-	0	0	0	0	0
S_6	0	0	0	0	0	-	0	0	**1**	0
S_7	0	0	0	0	0	**2**	-	0	0	0
S_8	0	0	0	**1**	**1**	0	0	-	0	0
S_9	0	0	0	0	0	0	0	0	-	0
S_{10}	**1**	0	0	0	0	**2**	0	0	**1**	-

Database Hacking: A solution to this attack is to check if the features are exactly (or extremely highly) the same as the stored features. In that case, system realizes that the provided features are stolen from the database, and it will deny access. This is based on the chaotic nature of the brain signals, where it is almost impossible that another signal has exactly the same features as the stored features of the user, although the two features can belong to the same class. From our experiments we see that features collected from the same person at different times have non-zero error with respect to the features used during training, hence supporting our claim.

Communication Snooping: A solution to this type of attack is to update the user signature in the database with the latest features that were granted access. In this situation, same as the solution in database hack attack, when the attacker provides the features to the system because its same as the stored features, it will not get access. Another solution is to keep all the feature sets which the system has accepted. In this case, as the attacker tries to get access using the stolen features, it will get checked against all the feature sets in the database and access will be denied.

4.2 Novel Presentation Attack Against Brain-Based Authentication Systems

Security systems using brain signals or Electroencephalogram (EEG), attempt to exploit chaotic nature of brain signals and their individuality to derive security primitives that are hard to reproduce. In this sense, the signal features are extracted to train a Machine Learning (ML) algorithm for classification. However, although brain signals are chaotic, feature extraction process might reduce the chaos rendering features in a way that they can be generated. Besides, even if features are chaotic, ML techniques might classify them in such a manner that an element in a particular class becomes easy to generate. We perform entropy analysis on common features used in EEG-based security systems to estimate their information content, which is used to propose a novel technique for EEG signal generation in feature domain instead of time domain. These generated signals can potentially be used for spoofing attacks. We consider five types of feature extraction techniques and six classifiers found in recently proposed security systems, and analyze their vulnerability to spoofing attacks using generated EEG signals. The results show that the generation scheme can synthesize artificial signals to get classified as genuine brain signals by ML algorithms.

4.2.1 Introduction, System and Attack Model

Electroencephalogram (EEG) monitoring using wireless, easy-to-wear, non-invasive and low cost sensors that connect to desktop and mobile phones, has enabled ubiquitous brain sensing and associated cognitive applications (Oskooyee *et al.* (2014); Pore *et al.* (2015); Sadeghi *et al.* (2016a,c)). Such seamless access to a subject's EEG signals, which contain information that are unique to her, nearly impossible to be covertly acquired, and chaotic over time, opens up opportunity for security applications. Indeed, several researchers have proposed EEG-based Security Systems

(ESS) for authentication of credentials, and identification of an entity (Sohankar *et al.* (2015b)).

The hypothesis in such systems is that the brain signals or features derived from them are chaotic, and vary over time in an unpredictable manner (even under same cognitive state or environmental conditions), but yet maintain individual characteristics, and hence have high information content or entropy (Shen *et al.* (2010)). This makes it difficult for an adversary to guess the original brain signal or generate a forged one to impersonate another person and successfully execute a *spoofing attack*. So, same as other biometric-based security systems, ESS uses Machine Learning (ML) techniques to find match between a subject EEG samples through classification. To build a high performance classifier, as seen in Fig. 4.6, feature extraction process reduces the dimension of the training data. But, the extracted features typically have lower entropy than time domain data. Loss of entropy in data features facilitates generating artificial features to break the security system, and opens up the system to a plethora of *spoofing attacks* (Sadeghi *et al.* (2016b)). ML-based generative models (Samanta (2011)) can be used to forge low entropy feature samples.

In previous studies, some attacks against biometric-based security systems are presented and studied. For instance, in (Chuang *et al.* (2013)), the possibility of impersonation attacks by knowing the cognitive state or secret pass-thought is proposed as an open problem. Then, in (Johnson *et al.* (2014)), they analyze their system against impersonation attacks. Also, in (Maiorana *et al.* (2013)), the vulnerability of EEG-based biometric system to hill-climbing attacks is analyzed, and possible countermeasures are proposed. However, research on systematic attacks by exploiting knowledge about the inner specifications (i.e. data and algorithms) of the system is lacking. In this paper, we present a novel type of spoofing attacks based on developing a Generative Model (GM) of feature domain data. Our experiment on complexity of EEG signals show that entropy of EEG signals in frequency domain is the lowest among both time and feature domains data. We develop a GM of frequency domain data using Adaptive Network Fuzzy Inference System (ANFIS) that shows relatively better performance as compared to other popular ML algorithms in modeling chaotic time-series (Samanta (2011)). The obtained signals from generated frequency samples can be exploited as forged data of a legitimate subject to penetrate a security system. The results of the attack simulation on ESS that use the most common feature extraction and classification methods, show this new strategy of synthesizing EEG signals can initiate serious attacks to ESS. We make the following contributions: 1) entropy analysis on EEG signals in time and frequency domains to evaluate the effort required by an adversary to guess the features used by security systems, 2) presenting a novel spoofing attack against ESS based on a novel strategy of generating EEG signals with features similar to genuine brain signals, and 3) comprehensive simulation of spoofing attacks by exploiting various forged signals including our generated signals to evaluate the robustness of ESS.

EEG-Based Security Systems

EEG-based security systems have three main phases: 1) *data collection*, 2) signal processing (i.e. *feature extraction* and *classification*), and 3) *decision making.* As seen in Fig. 4.6, first, sensors collect EEG from a subject who is asked to perform a specific cognitive task. For example, the subject might be asked to think of a specific word, imagine body movement or simply be in a relaxed state (Sadeghi and et. al (2011); Oskooyee *et al.* (2011)). Then, in the signal processing phase, features are extracted from the collected signals. The common feature extraction methods used in ESS include: Statistical Analysis (SA) such as mean of signals, Power Spectral Density (PSD), Auto Regressive (AR) coefficients, Fast Fourier Transform (FFT), and

Discrete Wavelet Transform (DWT). After feature extraction, a classifier maps each signal epoch to a specific class labeled for each subject. The classifier recognizes the unique characteristics of a subject's brain signals. ML techniques are widely used for EEG signal classification in security systems, such as Cosine Similarity Test (CST), Linear Discriminant Analysis (LDA), Naïve Bayes Classifier (NBC), Support Vector Machines (SVMs), k-Nearest Neighbors (kNNs), and Neural Networks (NNs) (Del Pozo-Banos *et al.* (2014)). Finally, in decision making phase, based on the classification results (also known as Matching Scores (MS)), the final decision about granting access or identifying the subject is made. Through these phases, ESS can provide three security services: 1) *registration*, 2) *authentication*, and 3) *identification*. In registration, we collect an epoch of EEG data from a subject, features are then extracted from the epoch, and stored in a database also denoted as *signature*. In authentication, the returning subject who wants to gain access as a registered subject, provides a claimed identity and EEG sample (*input signal*) typically shorter than the epoch in registration phase. Then, the feature set related to the claimed identity is compared with features of the collected data from the returning subject by applying ML techniques to grant access. In identification, the returning subject again provides input signal which is compared with all existing signatures in the system for a match.

Threat Model and Attack Scenarios

ESS spoofing attacks can be categorized in four main groups: 1) *impersonation*: mimicking a signal of a registered subject by performing a same cognitive task, 2) *replay attack*: reusing stolen signal of a registered subject, 3) *signal conversion*: altering available signal to reach a genuine signal (e.g. through hill-climbing), and 4) *signal synthesis*: generating an artificial signal based on available samples (Evans *et al.* (2013)). There are some effective counter-measures for the first three groups (Roberts

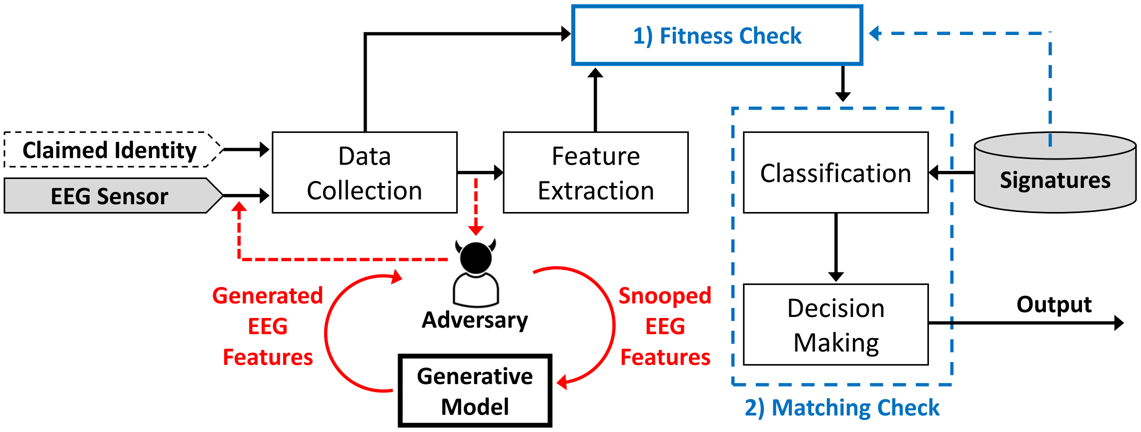


Figure 4.6: EEG-based Security System and Threat Model.

(2007); Gomez-Barrero *et al.* (2012)), while, synthesized signals cannot be detected easily. Our proposed attack method is based on signal synthesis, where an adversary with access to EEG data sample of a subject, derives data features, and develop a generative model of the relatively low entropy features. The generated features can be used to generate time domain data by inverting the feature extraction process. This step provides a signal that is different from the original signal, but has the intrinsic features that can be recognized as a registered entry by the classifier, in spoofing attacks. In our threat model (Fig. 4.6), we assume that an adversary has access to a registered subject EEG sample (*first assumption*). The adversary can prepare the following attacks:

1. *Replay attack:* Replaying the original snooped time domain signal sample.

2. *Spoofing with Temporal Noise (STN):* Adding white noise to the snooped time domain signal to prevent detection from similarity check.

3. *Spoofing with Spectral Noise (SSN):* Adding white noise to derived feature vectors from the snooped time domain signal sample, and returning it to time domain.

4. *Spoofing with Temporal Synthesis (STS):* Training ML-based GMs using the snooped time domain data, and generating a time-series potentially similar to the original signals.

5. *Spoofing with Spectral Synthesis (SSS):* Training ML-based GMs using derived features from snooped time domain data, and generating a time-series potentially having similar features as the original signals.

These attacks are highly dependent on the adversary's capability to generate either the time domain brain signals or the unique features of a registered subject. Hence, the success of these attacks are solely dependent on the chaotic properties or entropy of the brain signals. As the *second assumption*, the adversary has access to the ESS output.

EEG Signal Information content

For data analysis, to measure the chaos in a signal, we calculate the Shannon Entropy (ShEn) (Kannathal *et al.* (2005)) on the dataset described in Sec. 4.2.3. Signals with higher entropy are more chaotic and harder to generate. To calculate ShEn, a signal amplitude range is divided into fixed number of bins (ϵ) starting from minimum to maximum amplitude. Then, histogram (ρ_ϵ) for signal amplitude in different bins is obtained by dividing the number of samples in each bin by the total number of samples. Finally, the entropy is calculated as $ShEn = -\sum_\epsilon \rho_\epsilon \log \rho_\epsilon$. In time domain, raw signals are divided into 1 s epochs (considering the 160 Hz sampling rate, there is 160 samples per second) and entropy is calculated for each epoch, and then averaged over all epochs. For feature domain data, for instance, FFT is applied on each epoch to extract α band (8-13 Hz) feature vectors with length 6, and then the vector entropy is calculated. Similarly, DWT, AR, PSD, and SA feature vectors are derived with

Table 4.4: EEG Features Entropy Measurement.

Feature	Raw Data	FFT	DWT	AR	PSD	SA
Entropy	6.11	2.58	3.46	2.98	2.72	3.07

lengths 11, 10, 6 and 6, respectively (more details about the feature extraction are in Sec. 4.2.3). Tab. 4.4 shows, feature extraction reduces the signal entropy and its predictability. In our attack, we apply a generative model on feature domain data for accurate data generation.

4.2.2 EEG generative model

The EEG generative model is a ML system, which learns data parameters of a training dataset, and then uses those parameters to generate EEG signals. We use ANFIS (Samanta (2011)) to develop a generative model of EEG signal. In our model, to estimate the next p samples, also referred to as the *estimation window*, previous data samples (*training window*) are provided as inputs (x_{t-24}, x_{t-18}, x_{t-12}, and x_{t-6}, where t indicates the current time stamp) to the ANFIS model to calibrate it's parameters through training phase. In the training phase, the parameters of ANFIS are found in such a way that the error between the model output (x'_t) and the target output (x_t) falls below a threshold, or fixed number of optimization iterations are reached. The trained ANFIS can estimate the future value of data sample (x_{t+1}) based on previous values (x_{t-23}, x_{t-17}, x_{t-11}, and x_{t-5}). Subsequently, for estimating more future values up to x_{t+p}, we use the estimated value for x_{t+1} as a part of input to estimate x_{t+7}, and so forth for estimating x_{t+8} to x_{t+p}, we use the previous estimated values.

In our proposed approach for generating EEG time-series (Fig. 4.8), instead of generating EEG signals in time domain, we develop the GM in frequency domain. As

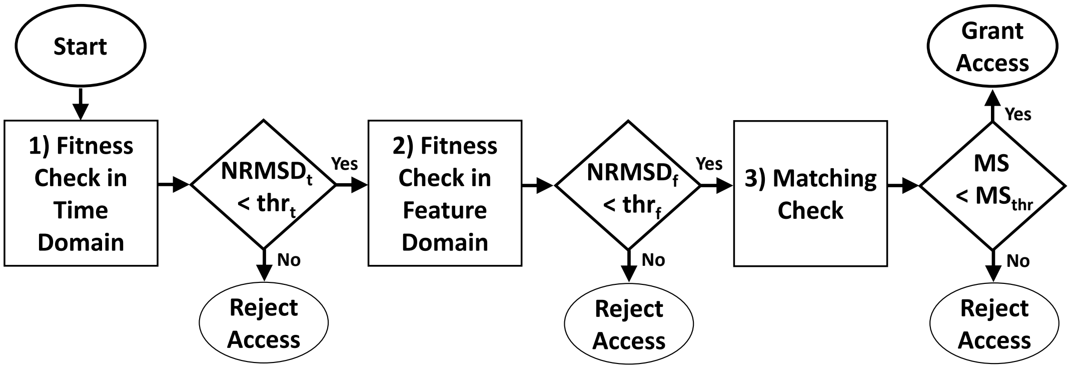


Figure 4.7: Detailed Mechanism of Fitness and Matching Checks.

mentioned in Sec. 4.2.1, entropy in frequency domain is much less than that of time domain and other features domains, which means less estimation error and more accurate model. In this sense, first, a time-series signal is divided into same size epochs, and FFT is applied on each epoch to derive the *spectrogram* of EEG signals (captured with *2N* sampling frequency) from *1* to N Hz. Real and imaginary parts are separated to form two individual vectors, and ANFIS is applied on each of them for estimating next points. The estimated points of two vectors are used as inputs to Inverse Fast Fourier Transform (IFFT) function. The output of the IFFT is the estimated signal in time domain. The GM is developed for each of the N frequency vectors, and the combination of the one step ahead estimation of these N GMs in frequency domain is equivalent to one second ahead estimation in time domain. In time domain estimation, *2N* step ahead should be estimated to have 1 s of data, while in our method N one step ahead estimation suffices. As the estimation window increases, the error will also increase. So, our method is more accurate because of using lower number of estimated steps ahead.

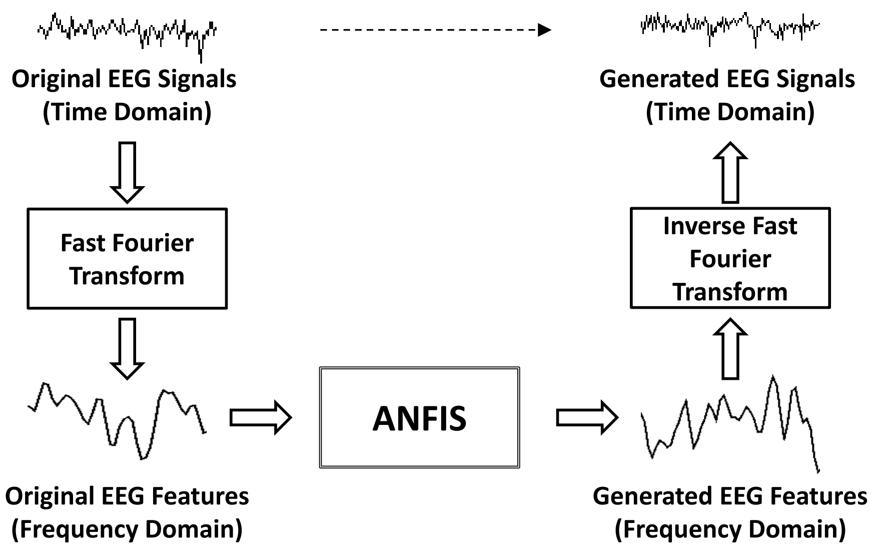


Figure 4.8: EEG Generation using Frequency Domain Signals.

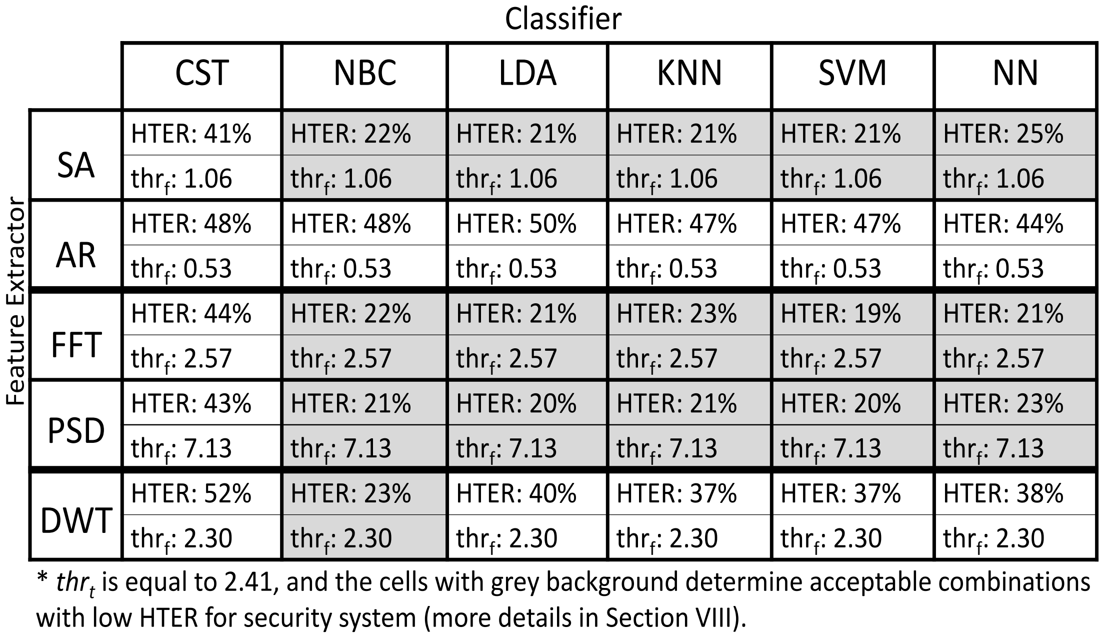
4.2.3 Security System Simulation

In system simulation, to grant access to the system, three checkpoints should be passed as shown in Fig. 4.7. In the first two checkpoints, similarity between current signal and existing signals in the system is measured. If the pair of signals were recognized as similar signals in either time or frequency domain, there would be a doubt on exploiting stolen data from previously used signals and the access request is rejected. At the third checkpoint, input signal goes through a classifier to see if it can be classified in a same class as a registered subject. In this section, we elaborate the system setup, simulation, and performance.

Dataset, Preprocessing, and Feature Extraction

In our experiment, we use raw EEG signals from 106 subjects with sampling rate of 160 Hz (Schalk *et al.* (2004); Goldberger and et. al (2000)). We choose channel "C3" signals from three 2-min sessions of opening and closing left or right fist for each of 106 subjects. The signal in each session is divided into four segments. The

Table 4.5: Performance of The Simulated Security System.



	Classifier					
Feature Extractor	CST	NBC	LDA	KNN	SVM	NN
SA	HTER: 41%	HTER: 22%	HTER: 21%	HTER: 21%	HTER: 21%	HTER: 25%
	thr_f: 1.06	thr_f: 1.06	thr_f: 1.06	thr_f: 1.06	thr_f: 1.06	thr_f: 1.06
AR	HTER: 48%	HTER: 48%	HTER: 50%	HTER: 47%	HTER: 47%	HTER: 44%
	thr_f: 0.53	thr_f: 0.53	thr_f: 0.53	thr_f: 0.53	thr_f: 0.53	thr_f: 0.53
FFT	HTER: 44%	HTER: 22%	HTER: 21%	HTER: 23%	HTER: 19%	HTER: 21%
	thr_f: 2.57	thr_f: 2.57	thr_f: 2.57	thr_f: 2.57	thr_f: 2.57	thr_f: 2.57
PSD	HTER: 43%	HTER: 21%	HTER: 20%	HTER: 21%	HTER: 20%	HTER: 23%
	thr_f: 7.13	thr_f: 7.13	thr_f: 7.13	thr_f: 7.13	thr_f: 7.13	thr_f: 7.13
DWT	HTER: 52%	HTER: 23%	HTER: 40%	HTER: 37%	HTER: 37%	HTER: 38%
	thr_f: 2.30	thr_f: 2.30	thr_f: 2.30	thr_f: 2.30	thr_f: 2.30	thr_f: 2.30

* thr_t is equal to 2.41, and the cells with grey background determine acceptable combinations with low HTER for security system (more details in Section VIII).

first segments (i.e. the first 30 s of each session) are used for signatures in registration (Sec. 4.2.1). The second and third segments are used for training the classifiers and deriving classification thresholds. The forth segments are used as input signal to test the security system. For preprocessing, we normalize the raw EEG with zero-mean and unit-variance methods. Also, a 5^{th} order Butterworth band-pass filter is applied on the signal in 8-13 Hz (a.k.a. α frequency band). For feature extraction, we use SA, AR, FFT, PSD, and DWT. For our SA feature vector, we choose mean, skewness, kurtosis, standard deviation, entropy, and range of a given signal. Also, 10^{th} order AR coefficients are extracted for AR features. In FFT, α band is used as frequency domain features. For PSD features, power spectral density is calculated on α band. And, 4^{th} level approximation coefficients of DWT are extracted as DWT features.

Table 4.6: Vulnerability Test Results for STN Simulation.

		Classifier					
		CST	NBC	LDA	KNN	SVM	NN
Feature Extractor	SA	SR: 69% (-2)	SR: 0% (-1)	SR: 2% (-1)	SR: 5% (-1)	SR: 4% (-1)	SR: 6% (-1)
		NRMSD_f: 1.59	NRMSD_f: 1.39	NRMSD_f: 1.39	NRMSD_f: 1.39	NRMSD_f: 1.39	NRMSD_f: 1.39
		NRMSD_t: 2.39	NRMSD_t: 2.26	NRMSD_t: 2.26	NRMSD_t: 2.26	NRMSD_t: 2.26	NRMSD_t: 2.26
	AR	SR: 0% (-7)	SR: 0% (-7)	SR: 0% (-7)	SR: 0% (-7)	SR: 0% (-7)	SR: 0% (-7)
		NRMSD_f: 0.98	NRMSD_f: 0.98	NRMSD_f: 0.98	NRMSD_f: 0.98	NRMSD_f: 0.98	NRMSD_f: 0.98
		NRMSD_t: 3.44	NRMSD_t: 3.44	NRMSD_t: 3.44	NRMSD_t: 3.44	NRMSD_t: 3.44	NRMSD_t: 3.44
	FFT	SR: 26% (-5)	SR: 4% (-5)	SR: 4% (-5)	SR: 21% (-5)	SR: 5% (-6)	SR: 11% (-6)
		NRMSD_f: 1.85	NRMSD_f: 1.85	NRMSD_f: 1.85	NRMSD_f: 1.85	NRMSD_f: 1.90	NRMSD_f: 1.90
		NRMSD_t: 2.92	NRMSD_t: 2.92	NRMSD_t: 2.92	NRMSD_t: 2.92	NRMSD_t: 3.17	NRMSD_t: 3.17
	PSD	SR: 40% (-1)	SR: 15% (-1)	SR: 21% (-1)	SR: 30% (-1)	SR: 16% (-1)	SR: 21% (-1)
		NRMSD_f: 3.06	NRMSD_f: 3.06	NRMSD_f: 3.06	NRMSD_f: 3.06	NRMSD_f: 3.06	NRMSD_f: 3.06
		NRMSD_t: 2.26	NRMSD_t: 2.26	NRMSD_t: 2.26	NRMSD_t: 2.26	NRMSD_t: 2.26	NRMSD_t: 2.26
	DWT	SR: 51% (-5)	SR: 11% (-1)	SR: 26% (-1)	SR: 75% (-2)	SR: 20% (-1)	SR: 30% (-1)
		NRMSD_f: 5.48	NRMSD_f: 3.68	NRMSD_f: 3.68	NRMSD_f: 4.03	NRMSD_f: 3.68	NRMSD_f: 3.68
		NRMSD_t: 2.92	NRMSD_t: 2.26	NRMSD_t: 2.26	NRMSD_t: 2.39	NRMSD_t: 2.26	NRMSD_t: 2.26

* Values in parentheses are *SNR*s that give the highest *SR*.

Fitness Check

To block replay attack, input signals should be compared with previously used data. We use Normalized Root Mean Square Deviation (NRMSD) between two signals to check their similarity Samanta (2011). NRMSD changes from zero to infinity, where a zero value indicates two given signals are identical, and larger values are interpreted as more dissimilarity between signals. We assume if NRMSD between stored signals and current signal drops below a threshold, which is experimentally set by calculating NRMSD for original signals, the attempt is considered as a suspicious activity. A signal with relatively high NRMSD can bypass fitness check of system and attack other components (Fig. 4.7).

Training data is exploited to calculate fitness thresholds in time ($\mathbf{thr_t}$) and frequency ($\mathbf{thr_f}$) domains. We use the term $S_{i,j}^{(k)}$ to label each segment in the data schema, where i, j and k refer to session, segment, and subject number, respectively. The signature ($S_{sig}^{(k)}$) for each subject is made by concatenating, the first 10 s of $S_{1,1}^{(k)}$, the second 10 s of $S_{2,1}^{(k)}$, and the last 10 s of $S_{3,1}^{(k)}$, making a new time series with 30 s length. To compute thr_t for session i of subject k, the fitness between $S_{i,1}^{(k)}$ and all remaining signature and training segments is calculated. The average of the resulting eight fitness values is considered as the fitness threshold related to $S_{i,1}^{(k)}$, and same computation is applied for calculating threshold relevant to $S_{i,2}^{(k)}$ and $S_{i,3}^{(k)}$. Finally, the maximum threshold value is chosen as the fitness threshold for session i. The thr_t is set to be the average of all sessions minus three times standard deviation. To obtain threshold in frequency domain, a similar procedure is used, and results in five various thr_f for the features (i.e. SA, AR, FFT, PSD, and DWT) (Tab. 4.5). In time domain fitness checkpoint (Fig. 4.7), the NRMSD in time domain ($NRMSD_t$), between signature ($S_{sig}^{(k)}$) and test data ($S_{i,4}^{(k)}$) of the claimed identity (k) is calculated. If the fitness is less than thr_t, signal will be considered as manipulated version of stolen signals. In frequency domain fitness checkpoint, the NRMSD in feature domain ($NRMSD_f$), between signature ($S_{sig}^{(k)}$) and test data ($S_{i,4}^{(k)}$) of the claimed identity (k) is calculated. If the fitness is less than thr_f, signal will be rejected.

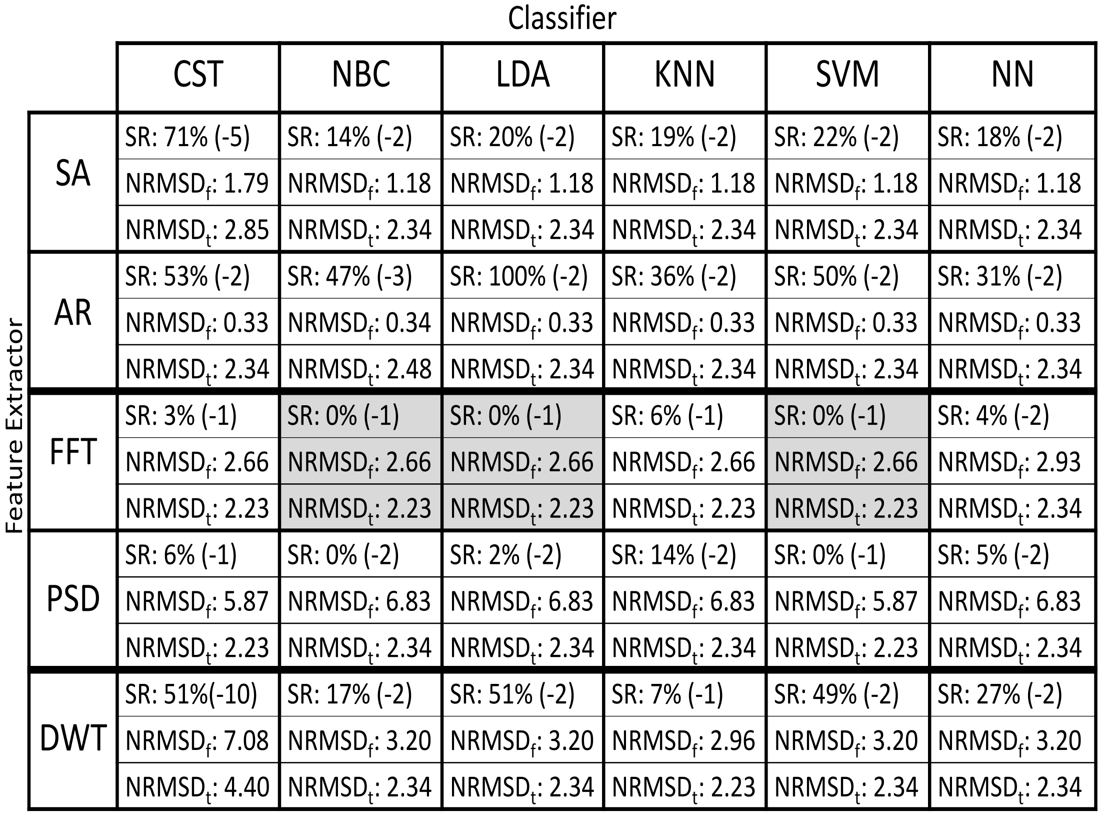
Matching Check

In the last checkpoint (Fig. 4.7), six classification techniques (i.e., CST, NBC, LDA, kNN, SVM, and NN) are exploited to check the matching possibility between signatures and test data. We use binary classification to distinguish between registered and non-registered subjects. To calculate Matching Score threshold ($\mathbf{MS_{thr}}$), the training data (Sec. 4.2.3) is used as input signals, and matching scores between signatures and input signals are computed using a given classifier. The obtained matching scores can be presented in a matrix with size 106 × 106. Each of the 106 diagonal values of index $(k,\, k)$ in the matrix is equal to the average of matching scores of these pairs: $(S^{(k)}_{1,2}, S^{(k)}_{sig})$, $(S^{(k)}_{1,3}, S^{(k)}_{sig})$, $(S^{(k)}_{2,2}, S^{(k)}_{sig})$, $(S^{(k)}_{2,3}, S^{(k)}_{sig})$, $(S^{(k)}_{3,2}, S^{(k)}_{sig})$, and $(S^{(k)}_{3,3}, S^{(k)}_{sig})$. To fill in the non-diagonal values of the matrix, for instance, to calculate the value of index (1, 2), training data of subject 2 should be classified with signature of subject 1. The matching score threshold of index (1, 2) is equal to average of six matching scores of the pairs: $(S^{(2)}_{1,2}, S^{(1)}_{sig})$, $(S^{(2)}_{1,3}, S^{(1)}_{sig})$, $(S^{(2)}_{2,2}, S^{(1)}_{sig})$, $(S^{(2)}_{2,3}, S^{(1)}_{sig})$, $(S^{(2)}_{3,2}, S^{(1)}_{sig})$, and $(S^{(2)}_{3,3}, S^{(1)}_{sig})$. Data related to same subject should be less separable, so it's expected that diagonal values be less than non-diagonal ones. We calculate the average of diagonal values and name it MS^{min}_{thr}, and also MS^{max}_{thr} is the average of non-diagonal values. To calculate the optimized value for MS_{thr}, we divide the interval between MS^{min}_{thr} and MS^{max}_{thr} into 100 parts, then one value is chosen for MS_{thr} among these 101 values, that minimizes the HTER of the system using the training data. Finally, test data is used as input signal to measure the performance of the ESS. A test input signal is accepted as a registered signal, if the matching score between it and the signature falls below MS_{thr}. Tab. 4.5 shows the HTER results using test data and based on the optimized value for MS_{thr}. Also, there may be more than one values for MS_{thr} that give us the same minimum HTER, in this case, we select the value closest to the mean value of MS^{min}_{thr} and MS^{max}_{thr}.

4.2.4 Spoofing Attack Simulation

For attack simulation, the system is evaluated under the five mentioned attack scenarios (Sec. 4.2.1). All possible combination of five feature extraction and six classification methods are tested based on each scenario (i.e. 30 cases). We calculate success rate (Sec. 4.2.1) of attacks for each scenario. In this sense, for a given subject

Table 4.7: Vulnerability Test Results for SSN Simulation.



		Classifier					
		CST	NBC	LDA	KNN	SVM	NN
Feature Extractor	SA	SR: 71% (-5) NRMSD_f: 1.79 NRMSD_t: 2.85	SR: 14% (-2) NRMSD_f: 1.18 NRMSD_t: 2.34	SR: 20% (-2) NRMSD_f: 1.18 NRMSD_t: 2.34	SR: 19% (-2) NRMSD_f: 1.18 NRMSD_t: 2.34	SR: 22% (-2) NRMSD_f: 1.18 NRMSD_t: 2.34	SR: 18% (-2) NRMSD_f: 1.18 NRMSD_t: 2.34
	AR	SR: 53% (-2) NRMSD_f: 0.33 NRMSD_t: 2.34	SR: 47% (-3) NRMSD_f: 0.34 NRMSD_t: 2.48	SR: 100% (-2) NRMSD_f: 0.33 NRMSD_t: 2.34	SR: 36% (-2) NRMSD_f: 0.33 NRMSD_t: 2.34	SR: 50% (-2) NRMSD_f: 0.33 NRMSD_t: 2.34	SR: 31% (-2) NRMSD_f: 0.33 NRMSD_t: 2.34
	FFT	SR: 3% (-1) NRMSD_f: 2.66 NRMSD_t: 2.23	SR: 0% (-1) NRMSD_f: 2.66 NRMSD_t: 2.23	SR: 0% (-1) NRMSD_f: 2.66 NRMSD_t: 2.23	SR: 6% (-1) NRMSD_f: 2.66 NRMSD_t: 2.23	SR: 0% (-1) NRMSD_f: 2.66 NRMSD_t: 2.23	SR: 4% (-2) NRMSD_f: 2.93 NRMSD_t: 2.34
	PSD	SR: 6% (-1) NRMSD_f: 5.87 NRMSD_t: 2.23	SR: 0% (-2) NRMSD_f: 6.83 NRMSD_t: 2.34	SR: 2% (-2) NRMSD_f: 6.83 NRMSD_t: 2.34	SR: 14% (-2) NRMSD_f: 6.83 NRMSD_t: 2.34	SR: 0% (-1) NRMSD_f: 5.87 NRMSD_t: 2.23	SR: 5% (-2) NRMSD_f: 6.83 NRMSD_t: 2.34
	DWT	SR: 51%(-10) NRMSD_f: 7.08 NRMSD_t: 4.40	SR: 17% (-2) NRMSD_f: 3.20 NRMSD_t: 2.34	SR: 51% (-2) NRMSD_f: 3.20 NRMSD_t: 2.34	SR: 7% (-1) NRMSD_f: 2.96 NRMSD_t: 2.23	SR: 49% (-2) NRMSD_f: 3.20 NRMSD_t: 2.34	SR: 27% (-2) NRMSD_f: 3.20 NRMSD_t: 2.34

k, we classify each test segment ($S_{i,4}^{(k)}$) with subjects' signature $S_{sig}^{(\xi)}$, $1 \leq \xi \leq 106$, and by comparing the matching scores and threshold MS_{thr}, we can calculate the security metrics (TA, FR, TR, and FA). At the end, the overall success rate is obtained by averaging over the three value sets of metrics related to each session:

Replay Attack

In this attack, we use original test data ($S_{i,4}^{(k)}$) from a given subject (k) to test the system vulnerability. In simulation, no changes are applied on input signals, so all the attempts are rejected in the fitness check phase, and success rate for all types of ESS is equal to zero.

Table 4.8: Vulnerability Test Results for STS Simulation

		Classifier					
		CST	NBC	LDA	KNN	SVM	NN
Feature Extractor	SA	SR: 13%	SR: 13%	SR: 29%	SR: 27%	SR: 31%	SR: 22%
		NRMSD$_f$: 0.68	NRMSD$_f$: 0.68	NRMSD$_f$: 0.68	NRMSD$_f$: 0.68	NRMSD$_f$: 0.68	NRMSD$_f$: 0.68
	AR	SR: 0%	SR: 0%	SR: 0%	SR: 0%	SR: 0%	SR: 0%
		NRMSD$_f$: 1.07	NRMSD$_f$: 1.07	NRMSD$_f$: 1.07	NRMSD$_f$: 1.07	NRMSD$_f$: 1.07	NRMSD$_f$: 1.07
	FFT	SR: 0%	SR: 6%	SR: 6%	SR: 2%	SR: 3%	SR: 11%
		NRMSD$_f$: 1.73	NRMSD$_f$: 1.73	NRMSD$_f$: 1.73	NRMSD$_f$: 1.73	NRMSD$_f$: 1.73	NRMSD$_f$: 1.73
	PSD	SR: 0%	SR: 15%	SR: 18%	SR: 8%	SR: 8%	SR: 16%
		NRMSD$_f$: 1.74	NRMSD$_f$: 1.74	NRMSD$_f$: 1.74	NRMSD$_f$: 1.74	NRMSD$_f$: 1.74	NRMSD$_f$: 1.74
	DWT	SR: 49%	SR: 54%	SR: 55%	SR: 9%	SR: 48%	SR: 51%
		NRMSD$_f$: 2.01	NRMSD$_f$: 2.01	NRMSD$_f$: 2.01	NRMSD$_f$: 2.01	NRMSD$_f$: 2.01	NRMSD$_f$: 2.01

* The average NRMSD in time domain (*$NRMSD_t$*) is equal to 1.63.

Spoofing with Temporal Noise (STN)

In STN, we add white Gaussian noise (with different Signal to Noise Ratio (SNR): $\{-10, -9, ..., -1, 1, ..., 10\}$) to $S_{i,4}^{(k)}$. The NRMSD between the original and synthesized signal is increased that makes it hard to be detected by fitness checkers, but also the classification results become worse which prohibits non-registered access. Tab. 4.6 demonstrates the best results for STN, where the SNR is set to a value that leads to maximum SR for a given feature extractor and classifier.

Spoofing with Spectral Noise (SSN)

In SSN, first we convert $S_{i,4}^{(k)}$ from time to frequency domain (8-13 Hz) by applying FFT. In the second step, white noise (similar to STN scenario) is added to the frequency domain signal. Finally, inverse FFT is used to convert back the altered signal to time domain for attack (Tab. 4.7). In each test case, the SNR is tuned to reach the highest success rate.

Spoofing with Temporal Synthesis (STS)

In STS (Tab. 4.8), ANFIS estimates one step ahead of $S_{i,4}^{(k)}$ in time domain. The number of training samples is $160 \times 3 = 480$, which is equivalent to 3 s recording of EEG. After each step, the training window shifts one sample ahead, and again ANFIS is trained for next estimation, and so on.

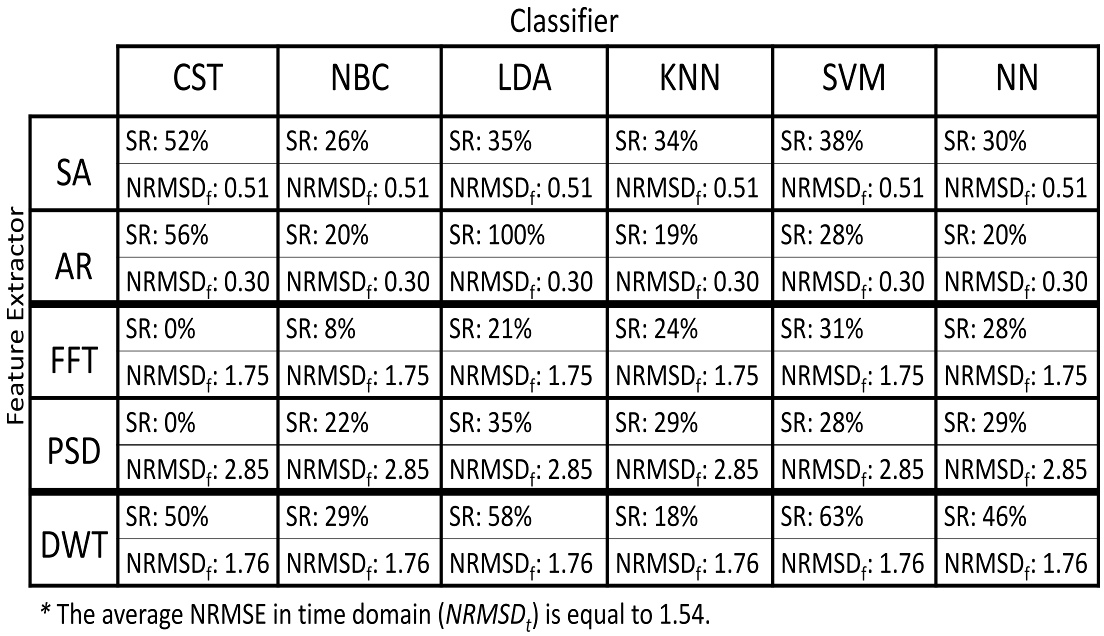
Spoofing with Spectral Synthesis (SSS)

In SSS, according to the discussed strategy in Sec. 4.2.2, FFT is used to convert $S_{i,4}^{(k)}$ from time to frequency domain. Then, one step ahead estimation is performed in frequency domain using ANFIS. Finally, IFFT converts back the estimated data into time domain. As seen in Tab. 4.9, although the NRMSD between the generated and original signals is high, but their classification results are close to each other.

Results Discussion

According to the evaluation results (Sec. 4.2.4), we select our *reference systems* according to the results in Tab. 4.5 (entries with grey background). A reference system is a combination of feature extractor and classifier that has HTER less than 25%. Then, in each remaining four attack simulations, some reference systems survive (the ones with success rate of 5% or less). As seen in Tab. 4.6, the pairs (SA & NBC), (SA & LDA), (SA & kNN), (SA & SVM), (FFT & NBC), (FFT & LDA), and (FFT & SVM) survive under STN attack. In SSN, (FFT & NBC), (FFT & LDA), and (FFT & SVM) remain robust so far. While, after STS attack, just (FFT & SVM) keeps its performance. Finally, in the last attack (SSS), none of the systems remain robust. The potential of passing through fitness and matching checkpoints by using

Table 4.9: Vulnerability Test Results for SSS simulation



	Classifier					
Feature Extractor	CST	NBC	LDA	KNN	SVM	NN
SA	SR: 52%	SR: 26%	SR: 35%	SR: 34%	SR: 38%	SR: 30%
	NRMSD$_f$: 0.51	NRMSD$_f$: 0.51	NRMSD$_f$: 0.51	NRMSD$_f$: 0.51	NRMSD$_f$: 0.51	NRMSD$_f$: 0.51
AR	SR: 56%	SR: 20%	SR: 100%	SR: 19%	SR: 28%	SR: 20%
	NRMSD$_f$: 0.30	NRMSD$_f$: 0.30	NRMSD$_f$: 0.30	NRMSD$_f$: 0.30	NRMSD$_f$: 0.30	NRMSD$_f$: 0.30
FFT	SR: 0%	SR: 8%	SR: 21%	SR: 24%	SR: 31%	SR: 28%
	NRMSD$_f$: 1.75	NRMSD$_f$: 1.75	NRMSD$_f$: 1.75	NRMSD$_f$: 1.75	NRMSD$_f$: 1.75	NRMSD$_f$: 1.75
PSD	SR: 0%	SR: 22%	SR: 35%	SR: 29%	SR: 28%	SR: 29%
	NRMSD$_f$: 2.85	NRMSD$_f$: 2.85	NRMSD$_f$: 2.85	NRMSD$_f$: 2.85	NRMSD$_f$: 2.85	NRMSD$_f$: 2.85
DWT	SR: 50%	SR: 29%	SR: 58%	SR: 18%	SR: 63%	SR: 46%
	NRMSD$_f$: 1.76	NRMSD$_f$: 1.76	NRMSD$_f$: 1.76	NRMSD$_f$: 1.76	NRMSD$_f$: 1.76	NRMSD$_f$: 1.76

* The average NRMSE in time domain (*NRMSD$_t$*) is equal to 1.54.

synthesized signals can increase FAR that poses serious threat against ESS.

4.3 Stimuli and Correlated Trait

The relation between this work (Sadeghi *et al.* (2016a)) and brain biometric systems comes from the work done on detecting impact of a stimuli on two correlated traits. Beside the brain signals (i.e. EEG) which was the primary trait, heart signals (i.e. ECG) were captured as secondary trait for detecting impact of a stimuli (i.e. stress). Heart signals was chosen as the secondary trait since it has correlation with brain signals for stress stimuli. The application was *nMovie*, which used brain signals from a viewer to determine individual mental state (i.e. nervous or not nervous), and blurs video frames based on the mental state. ECG side channel was used for two reasons: a) continuous validation of mental state, and b) adaptive tuning of EEG-based nervousness recognition algorithm parameters for improved accuracy. Stress stimuli can be detected in both brain and heart signals however on one hand brain responds in 250 ms but changes in heart rates are reflected in 3-4 s. On the other hand, accuracy of stress detection in brain signals is lower than heart signals. Basically, there is a trade-off between latency and accuracy of stimuli detection in the two trait.

The core idea of this technique (using stimuli and monitoring correlated trait) is applicable for detecting brain signal's timeliness (timely property of liveness). As discussed in section 2.1.5, timeliness detection is a challenge since normally there is no ground truth for it, which in return makes standalone methods infeasible. However, by adding context by either presenting stimuli, or capturing correlated trait, one can detect timeliness property.

In that work, two approaches for adding context was combined; stimuli was presented and its impact on both brain signal and the correlated heart signal was detected. While the technique allowed for robust detection of timeliness, it introduced several new challenges and costs. First, there is a need for an extra sensor (i.e. heart sensor) which increases cost and reduces usability. Second, extra delay (3-4 s) was introduced until the impact of stimuli showed up on correlated signal, which increases system latency and reduces usability. Third, processing of the correlated signal increases computation complexity. Forth, there is need for an stimuli to be presented to user, which not necessarily it has a universal and uniform impact on all users.

In summary, as that work showed timeliness detection is possible but comes with challenges such as extra sensor, increased latency and computation complexity, difficulty of finding universal stimuli, and reduced usability. Due to these challenges, this thesis focused on on detecting live property of brain signal. I focused on software solutions without the need for stimuli or correlated signals.

Chapter 5

CONCLUSION & FUTURE WORKS

Human inputs such as fingerprint touch, voice, face, and physiological signals (heart and brain) are replacing the traditional inputs (keyboard and mouse) in wide range of applications. Such systems which fall under cyber-physical systems are utilized in different aspects of social life such as medical, transportation, communication, military, and security. Utilizing human traits for authentication and access control has become more prevalent in recent decades in personal (e.g. face recognition in smartphones), business (e.g. workplace check in with fingerprint) and government (e.g. border control based on face and fingerprint) domains. Biometric systems which authenticate valid users based on traits such fingerprint, face, iris, and voice have been extensively researched and commercial systems have been developed. Utilizing physiological signals and especially brain Electroencephalography (EEG) signals is gaining momentum in biometric field due characteristics such as inherent inaccessibility (remote sensing is not possible), high entropy and chaotic nature, private (no signal traces left from everyday activities) and permanent (having brain until very late stages of life).

This thesis, proposes *E-BIAS*, EEG-based Authentication and Identification System, a brain-mobile security system which makes contributions in three directions. First, it provides high performance on signals with shorter length collected by commercial sensors and processed with lightweight models to meet computation/energy capacity of mobile devices. Second, to evaluate system's robustness a novel presentation attack was designed which challenged the literature's presumption of intrinsic liveness property for brain signals. Third, to bridge the gap, for the I formulated and studied brain liveness problem for the first time and proposed two solution approaches (*model-aware* & *model-agnostic*) to ensure liveness and enhance robustness against presentation attacks. Under each of two solution approaches, several methods were suggested and evaluated against both *synthetic* and *manipulative* class of attacks (total of 43 different attack vectors). Methods in both model-aware and model-agnostic approaches were successful in achieving error rate of zero 0%. More importantly, such error rates were reached in face of *unseen* attacks which provides evidence of the generalization potentials of the proposed solution approaches and methods. Generalization to new attacks is a significant factor in robustness of defense mechanisms since during design time there exist no knowledge about novel attacks that will surface in future. The search space for attacks is astronomically large and cannot be brute-forced using all the current and future computation power at human disposal. In the most simple case, if EEG signal is recorded using low sampling rate of 160Hz and each sample can take only 100 integer values in range of -50 to 50 micro-volts, one second of EEG signal can have $100^{1}60$ or $10^{3}20$ cases where the current supercomputers computation power is less than $10^{2}0$ FLOPS (floating point operations per second). So regardless of the number of attack vectors that one utilizes during design phase, new attacks can and will emerge in future. Hence, there should be special focus on how well defense methods can generalize against unseen attacks beside their robustness against the set of attacks used during design phase. Therefore, I suggested an adversarial work-flow to facilitate attack and defense cycles to allow for enhanced generalization capacity for domains in which decision-making process is non-deterministic such as cyber-physical systems (e.g. biometric/medical monitoring, autonomous machines, etc.). I utilized this work flow for brain liveness problem and was able to iteratively improve quality of both the designed attacks and the proposed liveness detection methods.

Moreover in this research, I systematically studied the state-of-the-art in biometric liveness detection methods for different traits and provided a taxonomy. While liveness methods were only evaluated based on their performance against attacks, I proposed a comprehensive evaluation criteria which also encompasses *usability* and *cost* factors and studies performance under several threat models instead of one. Based on the designed metric, the state-of-the-art in biometric liveness detection methods was evaluated and their trade-offs was discussed. Furthermore, I argued that liveness detection methods should shift from biological traits toward cognitive and intelligence capabilities of humans and proposed novel methods in this regard.

In the end, the proposed adversarial work-flow proved effective in application for brain liveness problem and it should be utilized and improved for the current and upcoming problems in ensuring source of inputs from physical world. Liveness problem will be of significant important for social life in upcoming decades due to increasing power of machines in automated behavior and content generation. With advances in Artificial Intelligence (AI), all different types of content/data can possibly be forged, and there is a need for liveness checking in all domains (e.g. text, audio, video, signal, art, etc.) to distinguish human generated works from those of machine. Already there exists works which perform well in generating different types of video (net, 2018d; Güera and Delp, 2018), text (Brown *et al.*, 2020), music (Dhariwal *et al.*, 2020) and art (Foster, 2019).

5.1 Future Research Direction

Building upon limitations of this thesis, I will discuss research directions to provide a road map for future studies.

The proposed E-BIAS(EEG-based Identification and Authentication System) focuses mostly on authentication task and less to identification task which is also not much researched in literature either. Brain biometric system should also decrease their latency to at least less than one second if they were to compete with the real-world fingerprint and face recognition systems. This means the length of the brain signal in hand would significantly reduce and new processing methods should be designed to allow for extremely low error rates in order of 10^{-5}. Furthermore, there is a need for gathering large public datasets of brain signals (with minimum thousands of subjects) and adversarial signals (different classes and techniques of attacks) to allow for benchmarking. Ideally, authentication should happen on the go without the need for a mental task (e.g. imagined hand movement) for user. Finally, permanency of the features that are extracted from brain signals through time should be studied comprehensively to allow for users to be verified for long period of times without the need for recalibration and another round of user registration.

Beside ensuring that incoming input trait has been sensed from a live human being, the *live* property (focus of this work), *timely* property should also be verified. In timeliness, system checks if input is been sensed at the current point in time and is not a replay of previous recording. Utilizing context through a challenge and response can help with ensuring timely property of the input. In a complete system, both properties should be tested to limit attack as much as possible.

The proposed adversarial work flow should be utilized in other biometric and cyber-physical systems (such as autonomous vehicles, smart homes, and robotics) so that its shortcomings and trade-offs are evaluated. Based on that, it can be improved and enhanced to be effective in systems with different set of requirements. This work used brain EEG signals as input for the workflow which is a time-series and a one-dimensional data. The framework should be adjusted and tuned for higher dimensional data (e.g. image and video) to scale its functionality.

The proposed evaluation criteria was utilized in a qualitative manner to evaluate liveness detection methods. Designing quantitative-based metrics would allow for more precise comparison between methods. Furthermore, liveness detection and authentication methods should be designed in concurrently and in respect to each other to provide increased robustness against presentation attacks.

In conjunction with liveness test where machines try to discern between a human and a machine, new methods should be designed to enable humans to effectively discern between the two (Turing test). This is of utmost importance since artificial intelligence advances has already allowed for machine which can generate artificial high quality contents (image, voice, video, text and art) that fool human observer. Furthermore, the ongoing migration to remote life style (accelerated by the COVID-19 global pandemic) asks for methods which can ensure the integrity of digital communications (e.g. online meetings, interviews, and exams) ans shared contents (e.g. image, voice and video). This is necessary to help with upholding social trust in everyday interactions.

REFERENCES

"2006 nist speaker recognition evaluation training set", *https* : *//catalog.ldc.upenn.edu/LDC*2011*S*09, accessed: 2021-06-16 (2006).

"Scientists grow mini-kidney in lab - uq news - the university of queensland, australia", `https://www.uq.edu.au/news/article/2015/10/scientists-grow-mini-kidney-lab`, (Accessed on 07/26/2021) (2013).

"Scientists coax stem cells to form 3-d mini lungs", `https://www.uofmhealth.org/news/archive/201503/scientists-coax-stem-cells-form-3-d-` (Accessed on 07/26/2021) (2015).

Information technology — Biometric presentation attack detection — Part 1:Framework, ISO/IEC 30107-1 (International Oraganization for Standardization (ISO), 2016).

"Deepfacelab is a tool that utilizes machine learning to replace faces in videos.", URL `https://github.com/iperov/DeepFaceLab`, [Online; accessed 27-June-2019] (2018a).

"Deepfakes and why the future of porn is terrifying", URL `https://www.highsnobiety.com/p/what-are-deepfakes-ai-porn/`, [Online; accessed 27-June-2019] (2018b).

"high-resolution (e.g. 2048x1024) photorealistic video-to-video translation", URL `https://github.com/NVIDIA/vid2vid`, [Online; accessed 27-June-2019] (2018c).

"You won't believe what obama says in this video!", URL `https://www.youtube.com/watch?v=cQ54GDm1eL0`, [Online; accessed 27-June-2019] (2018d).

"Color illusions", URL `http://brainden.com/color-illusions.htm`, [Online; accessed 27-June-2019] (2019a).

"Fakeapp", URL `https://www.fakeapp.org/`, [Offline; accessed 27-June-2019] (2019b).

"Free online barcode generator", URL `https://www.barcodesinc.com/generator/index.php`, [Online; accessed 27-June-2019] (2019c).

"'imagine this...' (2019) mark zuckerberg reveals the truth about facebook and who really owns the future", URL `https://www.instagram.com/p/ByaVigGFP2U/`, [Online; accessed 27-June-2019] (2019d).

"Myfakeapp", URL `https://bitbucket.org/radeksissues/myfakeapp/src/master/`, [Online; accessed 27-June-2019] (2019e).

"The superpower you always wanted", URL `https://twitter.com/deepnudeapp`, [Offline; accessed 27-June-2019] (2019f).

"Szondi test with pictures that will reveal your deepest hidden self", URL https://www.learning-mind.com/szondi-test-with-pictures-that-will-reveal-your-deepe [Online; accessed 27-June-2019] (2019g).

"Brain death", https://www.nhs.uk/conditions/brain-death/, accessed: 2021-05-01 (2021).

"Machine learning challenge winning solutions", https://github.com/dmlc/xgboost/tree/master/demo, (Accessed on 07/23/2021) (2021).

Abdulkader, S. N., A. Atia and M.-S. M. Mostafa, "Authentication systems: Principles and threats", Computer and Information Science **8**, 3, 155 (2015).

Abdullah, M. K., K. S. Subari, J. L. C. Loong and N. N. Ahmad, "Analysis of effective channel placement for an eeg-based biometric system", in "Biomedical Engineering and Sciences, IEEE EMBS Conference on", pp. 303–306 (2010).

Abhyankar, A. S. and S. C. Schuckers, "A wavelet-based approach to detecting liveness in fingerprint scanners", in "Biometric Technology for Human Identification", vol. 5404, pp. 278–286 (International Society for Optics and Photonics, 2004).

Abo-Zahhad, M., S. M. Ahmed and S. N. Abbas, "A new multi-level approach to eeg based human authentication using eye blinking", Pattern Recognition Letters **82**, 216–225 (2016).

Aides, A., D. David and H. Aronowitz, "Robust audiovisual liveness detection for biometric authentication using deep joint embedding and dynamic time warping", in "2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)", pp. 3026–3030 (IEEE, 2018).

Akhtar, Z., C. Micheloni and G. L. Foresti, "Biometric liveness detection: Challenges and research opportunities", IEEE Security & Privacy **13**, 5, 63–72 (2015).

Alegre, F., A. Janicki and N. Evans, "Re-assessing the threat of replay spoofing attacks against automatic speaker verification", in "2014 International Conference of the Biometrics Special Interest Group (BIOSIG)", pp. 1–6 (IEEE, 2014).

Almehmadi, A. and K. El-Khatib, "The state of the art in electroencephalogram and access control", in "Communications and Information Technology (ICCIT), 2013 Third International Conference on", pp. 49–54 (IEEE, 2013).

Alsufyani, N., A. Ali, S. Hoque and F. Deravi, "Biometric presentation attack detection using gaze alignment", in "2018 IEEE 4th International Conference on Identity, Security, and Behavior Analysis (ISBA)", pp. 1–8 (IEEE, 2018).

Anjos, A. and S. Marcel, "Counter-measures to photo attacks in face recognition: a public database and a baseline", in "2011 international joint conference on Biometrics (IJCB)", pp. 1–7 (IEEE, 2011).

Antonelli, A., R. Cappelli, D. Maio and D. Maltoni, “Fake finger detection by skin distortion analysis”, IEEE Transactions on Information Forensics and Security **1**, 3, 360–373 (2006).

Ashby, C., A. Bhatia, F. Tenore and J. Vogelstein, “Low-cost electroencephalogram (eeg) based authentication”, in “Neural Engineering, 5th International IEEE/EMBS Conference on”, pp. 442–445 (2011).

Aydoğdu, Ö., Z. Sadreddini and M. Ekinci, “A study on liveness analysis for palmprint recognition system”, in “2018 41st International Conference on Telecommunications and Signal Processing (TSP)”, pp. 1–4 (IEEE, 2018).

Banerjee, A., S. K. Gupta and K. K. Venkatasubramanian, “Pees: physiology-based end-to-end security for mhealth”, in “Proceedings of the 4th Conference on Wireless Health”, p. 2 (ACM, 2013).

Barra, S., “Design of a multi-biometric platform, based on physical traits and physiological measures: Face, iris, ear, ecg and eeg”, (2016).

Barreno, M., B. Nelson, R. Sears, A. D. Joseph and J. D. Tygar, “Can machine learning be secure?”, in “Proceedings of the 2006 ACM Symposium on Information, computer and communications security”, pp. 16–25 (ACM, 2006).

Benlamoudi, A., D. Samai, A. Ouafi, S. Bekhouche, A. Taleb-Ahmed and A. Hadid, “Face spoofing detection using multi-level local phase quantization (ml-lpq)”, (2015).

Boulkenafet, Z., J. Komulainen, Z. Akhtar, A. Benlamoudi, D. Samai, S. E. Bekhouche, A. Ouafi, F. Dornaika, A. Taleb-Ahmed, L. Qin *et al.*, “A competition on generalized software-based face presentation attack detection in mobile scenarios”, in “2017 IEEE International Joint Conference on Biometrics (IJCB)”, pp. 688–696 (IEEE, 2017a).

Boulkenafet, Z., J. Komulainen, L. Li, X. Feng and A. Hadid, “Oulu-npu: A mobile face presentation attack database with real-world variations”, in “2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017)”, pp. 612–618 (IEEE, 2017b).

Brigham, K. and B. V. Kumar, “Subject identification from electroencephalogram (eeg) signals during imagined speech”, in “2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)”, pp. 1–8 (IEEE, 2010).

Brown, T. B., B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, “Language models are few-shot learners”, arXiv preprint arXiv:2005.14165 (2020).

Cai, H. and K. K. Venkatasubramanian, “Detecting signal injection attack-based morphological alterations of ecg measurements”, in “Distributed Computing in Sensor Systems (DCOSS), 2016 International Conference on”, pp. 127–135 (IEEE, 2016).

Campbell, A. and T. Choudhury, “From smart to cognitive phones”, Pervasive Computing, IEEE **11**, 3 (2012).

Campisi, P. and D. La Rocca, “Brain waves for automatic biometric-based user recognition”, IEEE transactions on information forensics and security **9**, 5, 782–800 (2014).

Campisi, P., G. Scarano, F. Babiloni, F. D. Fallani, S. Colonnese, E. Maiorana and L. Forastiere, “Brain waves based user recognition using the “eyes closed resting conditions” protocol”, in “2011 IEEE International Workshop on Information Forensics and Security”, pp. 1–6 (IEEE, 2011).

Carlini, N., P. Mishra, T. Vaidya, Y. Zhang, M. Sherr, C. Shields, D. Wagner and W. Zhou, “Hidden voice commands”, in “25th USENIX Security Symposium (USENIX Security 16), Austin, TX”, (2016).

Chakka, M. M., A. Anjos, S. Marcel, R. Tronci, D. Muntoni, G. Fadda, M. Pili, N. Sirena, G. Murgia, M. Ristori *et al.*, “Competition on counter measures to 2-d facial spoofing attacks”, in “2011 International Joint Conference on Biometrics (IJCB)”, pp. 1–6 (IEEE, 2011).

Chan, P. P., W. Liu, D. Chen, D. S. Yeung, F. Zhang, X. Wang and C.-C. Hsu, “Face liveness detection using a flash against 2d spoofing attack”, IEEE Transactions on Information Forensics and Security **13**, 2, 521–534 (2017).

Chingovska, I., A. Anjos and S. Marcel, “On the effectiveness of local binary patterns in face anti-spoofing”, in “2012 BIOSIG-proceedings of the international conference of biometrics special interest group (BIOSIG)”, pp. 1–7 (IEEE, 2012).

Chingovska, I., A. Mohammadi, A. Anjos and S. Marcel, *Evaluation Methodologies for Biometric Presentation Attack Detection*, pp. 457–480 (Springer International Publishing, Cham, 2019), URL $https : //doi.org/10.1007/978-3-319-92627-8_20$.

Chingovska, I., J. Yang, Z. Lei, D. Yi, S. Z. Li, O. Kahm, C. Glaser, N. Damer, A. Kuijper, A. Nouak *et al.*, “The 2nd competition on counter measures to 2d face spoofing attacks”, in “2013 International Conference on Biometrics (ICB)”, pp. 1–6 (IEEE, 2013).

Chuang, J., H. Nguyen, C. Wang and B. Johnson, “I think, therefore i am: Usability and security of authentication using brainwaves”, in “Financial Cryptography and Data Security”, pp. 1–16 (Springer, 2013).

Coli, P., G. L. Marcialis and F. Roli, “Power spectrum-based fingerprint vitality detection”, in “2007 IEEE Workshop on Automatic Identification Advanced Technologies”, (2007).

Czajka, A., “Pupil dynamics for iris liveness detection”, IEEE Transactions on Information Forensics and Security **10**, 4, 726–735 (2015).

Czajka, A. and K. W. Bowyer, "Presentation attack detection for iris recognition: An assessment of the state of the art", ACM Computing Surveys (CSUR) **51**, 4, 86 (2018).

Damer, N., F. Boutros, A. M. Saladié, F. Kirchbuchner and A. Kuijper, "Realistic dreams: Cascaded enhancement of gan-generated images with an example in face morphing attacks", in "2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)", pp. 1–10 (2019).

Das, P., J. Mcfiratht, Z. Fang, A. Boyd, G. Jang, A. Mohammadi, S. Purnapatra, D. Yambay, S. Marcel, M. Trokielewicz *et al.*, "Iris liveness detection competition (livdet-iris)-the 2020 edition", in "2020 IEEE International Joint Conference on Biometrics (IJCB)", pp. 1–9 (IEEE, 2020).

Daugman, J., "Demodulation by complex-valued wavelets for stochastic pattern recognition", International Journal of Wavelets, Multiresolution and Information Processing **1**, 01, 1–17 (2003).

De Leon, P. L., M. Pucher, J. Yamagishi, I. Hernaez and I. Saratxaga, "Evaluation of speaker verification security and detection of hmm-based synthetic speech", IEEE Transactions on Audio, Speech, and Language Processing **20**, 8, 2280–2290 (2012).

Del Pozo-Banos, M., J. B. Alonso, J. R. Ticay-Rivas and C. M. Travieso, "Electroencephalogram subject identification: A review", Expert Systems with Applications **41**, 15, 6537–6554 (2014).

Derakhshani, R., S. A. Schuckers, L. A. Hornak and L. O'Gorman, "Determination of vitality from a non-invasive biomedical measurement for use in fingerprint scanners", Pattern recognition **36**, 2, 383–396 (2003).

Dhariwal, P., H. Jun, C. Payne, J. W. Kim, A. Radford and I. Sutskever, "Jukebox: A generative model for music", arXiv preprint arXiv:2005.00341 (2020).

Drahansky, M., "Experiments with skin resistance and temperature for liveness detection", in "2008 International Conference on Intelligent Information Hiding and Multimedia Signal Processing", pp. 1075–1079 (IEEE, 2008).

Duc, N. M. and B. Q. Minh, "Your face is not your password face authentication bypassing lenovo–asus–toshiba", Black Hat Briefings **4**, 158 (2009).

Eberz, S., N. Paoletti, M. Roeschlin, M. Kwiatkowska, I. Martinovic and A. Patané, "Broken hearted: How to attack ecg biometrics", (2017).

Evans, N. W., T. Kinnunen and J. Yamagishi, "Spoofing and countermeasures for automatic speaker verification.", in "Interspeech", pp. 925–929 (2013).

Exarchos, T. P., A. T. Tzallas, D. I. Fotiadis, S. Konitsiotis and S. Giannopoulos, "Eeg transient event detection and classification using association rules", IEEE Transactions on Information Technology in Biomedicine **10**, 3, 451–457 (2006).

Ferreira, A. L. S., L. C. de Miranda, E. E. C. de Miranda and S. G. Sakamoto, "A survey of interactive systems based on brain-computer interfaces", SBC Journal on Interactive Systems **4**, 1, 3–13 (2013).

Foster, D., *Generative deep learning: teaching machines to paint, write, compose, and play* (O'Reilly Media, 2019).

Fraschini, M., A. Hillebrand, M. Demuru, L. Didaci and G. L. Marcialis, "An eeg-based biometric system using eigenvector centrality in resting state brain networks", IEEE Signal Processing Letters **22**, 6, 666–670 (2014).

Galbally, J., F. Alonso-Fernandez, J. Fierrez and J. Ortega-Garcia, "A high performance fingerprint liveness detection method based on quality related features", Future Generation Computer Systems **28**, 1, 311–321 (2012a).

Galbally, J., J. Fierrez, F. Alonso-Fernandez and M. Martinez-Diaz, "Evaluation of direct attacks to fingerprint verification systems", Telecommunication Systems **47**, 3-4, 243–254 (2011).

Galbally, J., J. Ortiz-Lopez, J. Fierrez and J. Ortega-Garcia, "Iris liveness detection based on quality related features", in "2012 5th IAPR International Conference on Biometrics (ICB)", pp. 271–276 (IEEE, 2012b).

Garau, M., M. Fraschini, L. Didaci and G. L. Marcialis, "Experimental results on multi-modal fusion of eeg-based personal verification algorithms", in "2016 International Conference on Biometrics (ICB)", pp. 1–6 (IEEE, 2016).

Ghiani, L., G. L. Marcialis and F. Roli, "Fingerprint liveness detection by local phase quantization", in "Pattern Recognition (ICPR), 2012 21st International Conference on", pp. 537–540 (IEEE, 2012).

Ghiani, L., D. Yambay, V. Mura, S. Tocco, G. L. Marcialis, F. Roli and S. Schuckcrs, "Livdet 2013 fingerprint liveness detection competition 2013", in "2013 International Conference on Biometrics (ICB)", pp. 1–6 (IEEE, 2013).

Ghiani, L., D. A. Yambay, V. Mura, G. L. Marcialis, F. Roli and S. A. Schuckers, "Review of the fingerprint liveness detection (livdet) competition series: 2009 to 2015", Image and Vision Computing (2016).

Ghosh, R. and A. Negi, "Liveness detection using progressive eyelid tracking", US Patent 20,160,140,390 (2016).

Goldberger, A. L. and et. al, "Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals", Circulation pp. e215–e220 (2000).

Gomez-Barrero, M., J. Galbally, J. Fierrez and J. Ortega-Garcia, "Face verification put to test: A hill-climbing attack based on the uphill-simplex algorithm", in "Biometrics, 5th IAPR International Conference on", pp. 40–45 (IEEE, 2012).

Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, "Generative adversarial nets", Advances in neural information processing systems **27** (2014).

Gragnaniello, D., C. Sansone and L. Verdoliva, "Iris liveness detection for mobile devices based on local descriptors", Pattern Recognition Letters **57**, 81–87 (2015).

Gu, J., M. Y. Chen, E. Y. Du, K. Chan, E. Vural and S. Bandyopadhyay, "Image-based liveness detection for ultrasonic fingerprints", US Patent 20,160,070,968 (2016).

Güera, D. and E. J. Delp, "Deepfake video detection using recurrent neural networks", in "2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)", pp. 1–6 (IEEE, 2018).

Gui, Q., W. Yang, Z. Jin, M. V. Ruiz-Blondet and S. Laszlo, "A residual feature-based replay attack detection approach for brainprint biometric systems", in "2016 IEEE International Workshop on Information Forensics and Security (WIFS)", pp. 1–6 (IEEE, 2016).

Hadid, A., N. Evans, S. Marcel and J. Fierrez, "Biometrics systems under spoofing attack: an evaluation methodology and lessons learned", IEEE Signal Processing Magazine **32**, 5, 20–30 (2015).

He, C., X. Lv and Z. J. Wang, "Hashing the mar coefficients from eeg data for person authentication", in "Acoustics, Speech and Signal Processing. ICASSP. IEEE International Conference on", (2009).

Hema, C. and A. A. Osman, "Single trial analysis on eeg signatures to identify individuals", in "Signal Processing and Its Applications (CSPA), 2010 6th International Colloquium on", pp. 1–3 (IEEE, 2010).

Hema, C. R., M. Paulraj and H. Kaur, "Brain signatures: a modality for biometric authentication", in "Electronic Design, ICED International Conference on", pp. 1–4 (IEEE, 2008).

Hu, B., C. Mao, W. Campbell, P. Moore, L. Liu and G. Zhao, "A pervasive eeg-based biometric system", in "Proceedings of 2011 international workshop on Ubiquitous affective awareness and intelligent interaction", pp. 17–24 (ACM, 2011).

Humayed, A., J. Lin, F. Li and B. Luo, "Cyber-physical systems security—a survey", IEEE Internet of Things Journal **4**, 6, 1802–1831 (2017).

Ishikawa, Y., C. Yoshida, M. Takata and K. Joe, "Validation of eeg personal authentication with multi-channels and multi-tasks", (????).

Jain, A. K., K. Nandakumar and A. Ross, "50 years of biometric research: Accomplishments, challenges, and opportunities", Pattern Recognition Letters **79**, 80–105 (2016).

Jain, A. K., A. Ross and S. Pankanti, "Biometrics: a tool for information security", Information Forensics and Security, IEEE Transactions on **1**, 2, 125–143 (2006).

Jain, A. K., A. Ross and S. Prabhakar, "An introduction to biometric recognition", IEEE Transactions on circuits and systems for video technology **14**, 1, 4–20 (2004).

Jang, J.-S., "Anfis: adaptive-network-based fuzzy inference system", Systems, Man and Cybernetics, IEEE Transactions on **23**, 3, 665–685 (1993).

Jee, H.-K., S.-U. Jung and J.-H. Yoo, "Liveness detection for embedded face recognition system", International Journal of Biological and Medical Sciences **1**, 4, 235–238 (2006).

Johnson, B., T. Maillart and J. Chuang, "My thoughts are not your thoughts", in "Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication", pp. 1329–1338 (2014).

Jones, A. H., Z. B. Dizon and T. W. October, "Investigation of public perception of brain death using the internet", Chest **154**, 2, 286–292 (2018).

Kannathal, N., M. L. Choo, U. R. Acharya and P. Sadasivan, "Entropies for detection of epilepsy in EEG", Computer methods and programs in biomedicine pp. 187–194 (2005).

Khaitan, S. K. and J. D. McCalley, "Design techniques and applications of cyber-physical systems: A survey", IEEE Systems Journal **9**, 2, 350–365 (2015).

Khalifa, W., A. Salem, M. Roushdy and K. Revett, "A survey of eeg based user authentication schemes", in "Informatics and Systems, 8th International Conference on", pp. BIO–55 (IEEE, 2012).

Kinnunen, T., M. Sahidullah, H. Delgado, M. Todisco, N. Evans, J. Yamagishi and K. A. Lee, "The asvspoof 2017 challenge: Assessing the limits of replay spoofing attack detection", (2017).

Kiss, I. *et al.*, "Detector for recognizing the living character of a finger in a fingerprint recognizing apparatus", US Patent 6,175,641 (2001).

Kollreider, K., H. Fronthaler and J. Bigun, "Evaluating liveness by face images and the structure tensor", in "Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05)", pp. 75–80 (IEEE, 2005).

Komeili, M., N. Armanfard and D. Hatzinakos, "Liveness detection and automatic template updating using fusion of ecg and fingerprint", IEEE Transactions on Information Forensics and Security **13**, 7, 1810–1822 (2018).

Komogortsev, O. V., A. Karpov and C. D. Holland, "Attack of mechanical replicas: Liveness detection with eye movements", IEEE Transactions on Information Forensics and Security **10**, 4, 716–725 (2015).

Kong, X., W. Kong, Q. Fan, Q. Zhao and A. Cichocki, “Task-independent eeg identification via low-rank matrix decomposition”, in “2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)”, pp. 412–419 (IEEE, 2018).

Kose, N. and J.-L. Dugelay, “Reflectance analysis based countermeasure technique to detect face mask attacks”, in “Digital Signal Processing (DSP), 2013 18th International Conference on”, pp. 1–6 (IEEE, 2013).

Krishnan, A., T. Thomas, G. R. Nayar and S. S. Mohan, “Liveness detection in finger vein imaging device using plethysmographic signals”, in “International Conference on Intelligent Human Computer Interaction”, pp. 251–260 (Springer, 2018).

Lai, C.-I., A. Abad, K. Richmond, J. Yamagishi, N. Dehak and S. King, “Attentive filtering networks for audio replay attack detection”, in “ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)”, pp. 6316–6320 (IEEE, 2019).

Lapsley, P. D., J. A. Lee, D. F. Pare Jr and N. Hoffman, “Anti-fraud biometric scanner that accurately detects blood flow”, US Patent 5,737,439 (1998).

Lavrentyeva, G., S. Novoselov, E. Malykh, A. Kozlov, O. Kudashev and V. Shchemelinin, “Audio replay attack detection with deep learning frameworks.”, in “Interspeech”, pp. 82–86 (2017).

Lee, H. J., H. S. Kim and K. S. Park, “A study on the reproducibility of biometric authentication based on electroencephalogram (eeg)”, in “Neural Engineering, 2013 6th International IEEE/EMBS Conference on”, pp. 13–16 (2013).

Li, J., Y. Wang, T. Tan and A. K. Jain, “Live face detection based on the analysis of fourier spectra”, in “Defense and Security”, pp. 296–303 (International Society for Optics and Photonics, 2004).

Li, J.-W., “Eye blink detection based on multiple gabor response waves”, in “2008 International Conference on Machine Learning and Cybernetics”, vol. 5, pp. 2852–2856 (IEEE, 2008).

Li, Y., M.-C. Chang and S. Lyu, “In ictu oculi: Exposing ai created fake videos by detecting eye blinking”, in “2018 IEEE International Workshop on Information Forensics and Security (WIFS)”, pp. 1–7 (IEEE, 2018).

Li, Y. and S. Lyu, “Exposing deepfake videos by detecting face warping artifacts”, arXiv preprint arXiv:1811.00656 (2018).

Lin, Q., W. Li, X. Ning, X. Dong and P. Chen, “Liveness detection using texture and 3d structure analysis”, in “Chinese Conference on Biometric Recognition”, pp. 637–645 (Springer, 2016).

Liu, J., J. Yang, C. Wu and Y. Chen, “A liveness detection method based on blood volume pulse probing”, in “Chinese Conference on Biometric Recognition”, pp. 646–654 (Springer, 2016).

Ma, Z., J. Wang, P. Loskill, N. Huebsch, S. Koo, F. L. Svedlund, N. C. Marks, E. W. Hua, C. P. Grigoropoulos, B. R. Conklin *et al.*, "Self-organizing human cardiac microchambers mediated by geometric confinement", Nature communications **6**, 7413 (2015).

Maiorana, E. and P. Campisi, "Longitudinal evaluation of eeg-based biometric recognition", IEEE Transactions on Information Forensics and Security **13**, 5, 1123–1138 (2017).

Maiorana, E., G. E. Hine, D. La Rocca and P. Campisi, "On the vulnerability of an eeg-based biometric system to hill-climbing attacks algorithms' comparison and possible countermeasures", in "Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on", pp. 1–6 (IEEE, 2013).

Maiorana, E., D. La Rocca and P. Campisi, "Eigenbrains and eigentensorbrains: Parsimonious bases for eeg biometrics", Neurocomputing **171**, 638–648 (2016).

Marasco, E. and A. Ross, "A survey on antispoofing schemes for fingerprint recognition systems", ACM Computing Surveys (CSUR) **47**, 2, 28 (2015).

Marcel, S. and J. d. R. Millán, "Person authentication using brainwaves (eeg) and maximum a posteriori model adaptation", Pattern Analysis and Machine Intelligence, IEEE Transactions on **29**, 4, 743–752 (2007).

Marcel, S., M. S. Nixon, J. Fierrez and N. Evans, *Handbook of biometric anti-spoofing: Presentation attack detection* (Springer, 2019).

Marcialis, G. L., A. Lewicke, B. Tan, P. Coli, D. Grimberg, A. Congiu, A. Tidu, F. Roli and S. Schuckers, "First international fingerprint liveness detection competition—livdet 2009", in "International Conference on Image Analysis and Processing", pp. 12–23 (Springer, 2009).

Matsumoto, T., H. Matsumoto, K. Yamada and S. Hoshino, "Impact of artificial" gummy" fingers on fingerprint systems", in "Optical Security and Counterfeit Deterrence Techniques IV", vol. 4677, pp. 275–290 (International Society for Optics and Photonics, 2002).

Matthew, P., *Novel approaches to biometric security with an emphasis on liveness and coercion detection.*, Ph.D. thesis, Edge Hill University (2016).

Matthew, P. and M. Anderson, "Novel approaches to developing multimodal biometric systems with autonomic liveness detection characteristics", in "Intelligent Systems for Science and Information", pp. 121–138 (Springer, 2014a).

Matthew, P. and M. Anderson, "Novel categorisation techniques for liveness detection", in "Next Generation Mobile Apps, Services and Technologies (NGMAST), 2014 Eighth International Conference on", pp. 153–158 (IEEE, 2014b).

Matthew, P. and M. Anderson, "Developing coercion detection solutions for biometrie security", in "SAI Computing Conference (SAI), 2016", pp. 1123–1130 (IEEE, 2016).

Ming, Z., J. Chazalon, M. M. Luqman, M. Visani and J.-C. Burie, “Facelivenet: End-to-end networks combining face verification with interactive facial expression-based liveness detection”, in “2018 24th International Conference on Pattern Recognition (ICPR)”, pp. 3507–3512 (IEEE, 2018).

Mohammadi, G., P. Shoushtari, B. Molaee Ardekani and M. B. Shamsollahi, “Person identification by using ar model for eeg signals”, in “Proceeding of World Academy of Science, Engineering and Technology”, vol. 11, pp. 281–285 (2006).

Moon, Y. S., J. Chen, K. Chan, K. So and K. Woo, “Wavelet based fingerprint liveness detection”, Electronics Letters **41**, 20, 1112–1113 (2005).

Mulholland, T., T. McLaughlin and F. Benson, “Feedback control and quantification of the response of eeg alpha to visual stimulation”, Biofeedback and Self-regulation **1**, 4, 411–422 (1976).

Mura, V., L. Ghiani, G. L. Marcialis, F. Roli, D. A. Yambay and S. A. Schuckers, “Livdet 2015 fingerprint liveness detection competition 2015”, in “2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)”, pp. 1–6 (2015).

Mura, V., G. Orrù, R. Casula, A. Sibiriu, G. Loi, P. Tuveri, L. Ghiani and G. L. Marcialis, “Livdet 2017 fingerprint liveness detection competition 2017”, in “2018 International Conference on Biometrics (ICB)”, pp. 297–302 (IEEE, 2018).

Nagar, A., K. Nandakumar and A. K. Jain, “Multibiometric cryptosystems based on feature-level fusion”, IEEE transactions on information forensics and security **7**, 1, 255–268 (2012).

Nakanishi, I., S. Baba and C. Miyamoto, “Eeg based biometric authentication using new spectral features”, in “Intelligent Signal Processing and Communication Systems. International Symposium on”, pp. 651–654 (IEEE, 2009).

Nakanishi, I. and T. Maruoka, “Biometric authentication using evoked potentials stimulated by personal ultrasound”, in “2019 42nd International Conference on Telecommunications and Signal Processing (TSP)”, pp. 365–368 (IEEE, 2019).

Nicolas-Alonso, L. F. and J. Gomez-Gil, “Brain computer interfaces, a review”, Sensors **12**, 2, 1211–1279 (2012).

Niedermeyer, E. and F. L. da Silva, *Electroencephalography: basic principles, clinical applications, and related fields* (Lippincott Williams & Wilkins, 2005).

Nixon, K. A., V. Aimale and R. K. Rowe, “Spoof detection schemes”, in “Handbook of biometrics”, pp. 403–423 (Springer, 2008).

O’Shea, A., G. Lightbody, G. Boylan and A. Temko, “Neonatal seizure detection from raw multi-channel eeg using a fully convolutional architecture”, Neural Networks **123**, 12–25 (2020).

Oskooyee, K. S., A. Banerjee and S. K. S. Gupta, “Neuro movie theatre: A real-time internet-of-people based mobile application”, The 15th International Workshop on Mobile Computing Systems and Applications (2014).

Oskooyee, K. S., M. M. R. Kashani and A. Harounabadi, “Implementing a cognition cycle with words computation”, in “Computational Intelligence, Cognitive Algorithms, Mind, and Brain, 2011 IEEE Symposium on”, pp. 1–6 (IEEE, 2011).

Paranjape, R., J. Mahovsky, L. Benedicenti and Z. Koles, “The electroencephalogram as a biometric”, in “Electrical and Computer Engineering, 2001. Canadian Conference on”, vol. 2, pp. 1363–1366 (IEEE, 2001).

Parthasaradhi, S. T., R. Derakhshani, L. A. Hornak and S. A. Schuckers, “Time-series detection of perspiration as a liveness test in fingerprint devices”, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) **35**, 3, 335–343 (2005).

Patel, T. B. and H. A. Patil, “Combining evidences from mel cepstral, cochlear filter cepstral and instantaneous frequency features for detection of natural vs. spoofed speech”, in “Sixteenth annual conference of the international speech communication association”, (2015).

Pore, M., K. Sadeghi, V. Chakati, A. Banerjee and S. K. S. Gupta, “Enabling real-time collaborative brain-mobile interactive applications on volunteer mobile devices”, in “Proceedings of the 2nd International Workshop on Hot Topics in Wireless”, pp. 46–50 (ACM, 2015).

Poulos, M., M. Rangoussi, N. Alexandris, A. Evangelou *et al.*, “Person identification from the eeg using nonlinear signal classification”, Methods of information in Medicine **41**, 1, 64–75 (2002).

Raghavendra, R. and C. Busch, “Presentation attack detection algorithm for face and iris biometrics”, in “2014 22nd European Signal Processing Conference (EUSIPCO)”, pp. 1387–1391 (IEEE, 2014).

Raghavendra, R., K. B. Raja, S. Venkatesh, F. A. Cheikh and C. Busch, “On the vulnerability of extended multispectral face recognition systems towards presentation attacks”, in “2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)”, pp. 1–8 (IEEE, 2017).

Ramachandra, R. and C. Busch, “Presentation attack detection methods for face recognition systems: a comprehensive survey”, ACM Computing Surveys (CSUR) **50**, 1, 8 (2017).

Rattani, A. and A. Ross, “Automatic adaptation of fingerprint liveness detector to new spoof materials”, in “Biometrics (IJCB), 2014 IEEE International Joint Conference on”, pp. 1–8 (IEEE, 2014).

Raya-Rivera, A. M., D. Esquiliano, R. Fierro-Pastrana, E. López-Bayghen, P. Valencia, R. Ordorica-Flores, S. Soker, J. J. Yoo and A. Atala, “Tissue-engineered autologous vaginal organs in patients: a pilot cohort study”, The Lancet **384**, 9940, 329–336 (2014).

Reddy, P. V., A. Kumar, S. Rahman and T. S. Mundra, “A new antispoofing approach for biometric devices”, IEEE Transactions on Biomedical Circuits and Systems **2**, 4, 328–337 (2008).

Riera, A., A. Soria-Frisch, M. Caparrini, I. Cester and G. Ruffini, “1 multimodal physiological biometrics authentication”, Biometrics: Theory, Methods, and Applications pp. 461–482 (2009).

Riera, A., A. Soria-Frisch, M. Caparrini, C. Grau and G. Ruffini, “Unobtrusive biometric system based on electroencephalogram analysis”, EURASIP Journal on Advances in Signal Processing **2008**, 18 (2008).

Roberts, C., “Biometric attack vectors and defences”, Computers & Security **26**, 1, 14–25 (2007).

Rogmann, N. and M. Krieg, “Liveness detection in biometrics”, in “Biometrics Special Interest Group (BIOSIG), 2015 International Conference of the”, pp. 1–14 (IEEE, 2015).

Ross, A. and A. Jain, “Information fusion in biometrics”, Pattern recognition letters **24**, 13, 2115–2125 (2003).

Ruiz-Albacete, V., P. Tome-Gonzalez, F. Alonso-Fernandez, J. Galbally, J. Fierrez and J. Ortega-Garcia, “Direct attacks using fake images in iris verification”, in “European Workshop on Biometrics and Identity Management”, pp. 181–190 (Springer, 2008).

Sadeghi, K., A. Banerjee, J. Sohankar and S. K. Gupta, “Optimization of brain mobile interface applications using iot”, in “2016 IEEE 23rd International Conference on High Performance Computing (HiPC)”, pp. 32–41 (IEEE, 2016a).

Sadeghi, K., A. Banerjee, J. Sohankar and S. K. Gupta, “Toward parametric security analysis of machine learning based cyber forensic biometric systems”, in “Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on”, pp. 626–631 (IEEE, 2016b).

Sadeghi, K., A. Banerjee, J. Sohankar and S. K. S. Gupta, “Safedrive: An autonomous driver safety application in aware cities”, in “Pervasive Computing and Communication Workshops, IEEE Intl. Conf. on”, pp. 1–6 (2016c).

Sadeghi, K. and et. al, “The shadow of a real conscious mind”, in “Cognitive Informatics & Cognitive Computing, 2011 10th IEEE International Conference on”, pp. 412–418 (IEEE, 2011).

Sadeghi, K., J. Sohankar, A. Banerjee and S. K. Gupta, “A novel spoofing attack against electroencephalogram-based security systems”, in “Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld), 2017 Intl IEEE Conferences”, (IEEE, 2017).

Sahidullah, M., D. A. L. Thomsen, R. G. Hautamäki, T. Kinnunen, Z.-H. Tan, R. Parts and M. Pitkänen, “Robust voice liveness detection and speaker verification using throat microphones”, IEEE/ACM Transactions on Audio, Speech, and Language Processing **26**, 1, 44–56 (2018).

Samanta, B., “Prediction of chaotic time series using computational intelligence”, Expert Systems with Applications (2011).

Schalk, G., D. J. McFarland, T. Hinterberger, N. Birbaumer and J. R. Wolpaw, “BCI2000: a general-purpose brain-computer interface (BCI) system”, Biomedical Engineering, IEEE Transactions on **51**, 6, 1034–1043 (2004).

Sequeira, A. F., H. P. Oliveira, J. C. Monteiro, J. P. Monteiro and J. S. Cardoso, “Mobilive 2014-mobile iris liveness detection competition”, in “IEEE international joint conference on biometrics”, pp. 1–6 (IEEE, 2014).

Shang, J., S. Chen and J. Wut, “Srvoice: A robust sparse representation-based liveness detection system”, in “2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS)”, pp. 291–298 (IEEE, 2018).

Sharieh, S., A. Ferworn, V. Toronov and A. Abhari, “An ad-hoc network based framework for monitoring brain function”, in “Proceedings of the 11th communications and networking simulation symposium”, pp. 49–55 (ACM, 2008).

Shen, M., L. Lin, J. Chen and C. Q. Chang, “A prediction approach for multichannel EEG signals modeling using local wavelet SVM”, Instrumentation and Measurement, IEEE Transactions on (2010).

Shi, W., J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert and Z. Wang, “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network”, in “Proceedings of the IEEE conference on computer vision and pattern recognition”, pp. 1874–1883 (2016).

Shiota, S., F. Villavicencio, J. Yamagishi, N. Ono, I. Echizen and T. Matsui, “Voice liveness detection algorithms based on pop noise caused by human breath for automatic speaker verification”, in “Sixteenth Annual Conference of the International Speech Communication Association”, (2015).

Sohankar, J., K. Sadeghi, A. Banerjee and S. K. Gupta, “E-bias: A pervasive eeg-based identification and authentication system”, in “Proceedings of the 11th ACM Symposium on QoS and Security for Wireless and Mobile Networks”, pp. 165–172 (ACM, 2015a).

Sohankar, J., K. Sadeghi, A. Banerjee and S. K. S. Gupta, “E-BIAS: A pervasive EEG-based identification and authentication system”, Proceedings of the 10th ACM Symposium on QoS and Security for Wireless and Mobile Networks (2015b).

Sousedik, C. and C. Busch, “Presentation attack detection methods for fingerprint recognition systems: a survey”, Iet Biometrics **3**, 4, 219–233 (2014).

Sun, L., G. Pan, Z. Wu and S. Lao, “Blinking-based live face detection using conditional random fields”, in “International Conference on Biometrics”, pp. 252–260 (Springer, 2007).

Sundararajan, A., A. Pons and A. I. Sarwat, “A generic framework for eeg-based biometric authentication”, in “2015 12th International Conference on Information Technology-New Generations”, pp. 139–144 (IEEE, 2015).

Takahashi, K. and T. Murakami, “A measure of information gained through biometric systems”, Image and Vision Computing **32**, 12, 1194–1203 (2014).

Tan, D., “Brain-computer interfaces: applying our minds to human-computer interaction. informal proceedings what is the next generation of human-computer interaction?”, in “Workshop at CHI 2006”, (2006).

Tan, X., Y. Li, J. Liu and L. Jiang, “Face liveness detection from a single image with sparse low rank bilinear discriminative model”, in “European Conference on Computer Vision”, pp. 504–517 (Springer, 2010).

Taylor, J. M. and H. R. Sharif, “Security challenges and methods for protecting critical infrastructure cyber-physical systems”, in “Selected Topics in Mobile and Wireless Networking (MoWNeT), 2017 International Conference on”, pp. 1–6 (IEEE, 2017).

Thomas, K. P. and A. P. Vinod, “Biometric identification of persons using sample entropy features of eeg during rest state”, in “2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)”, pp. 003487–003492 (IEEE, 2016).

Todisco, M., X. Wang, V. Vestman, M. Sahidullah, H. Delgado, A. Nautsch, J. Yamagishi, N. Evans, T. H. Kinnunen and K. A. Lee, “ASVspoof 2019: Future Horizons in Spoofed and Fake Audio Detection”, in “Proc. Interspeech 2019”, pp. 1008–1012 (2019), URL `http://dx.doi.org/10.21437/Interspeech.2019-2249`.

Tome, P., R. Raghavendra, C. Busch, S. Tirunagari, N. Poh, B. Shekar, D. Gragnaniello, C. Sansone, L. Verdoliva and S. Marcel, “The 1st competition on counter measures to finger vein spoofing attacks”, in “2015 International Conference on Biometrics (ICB)”, pp. 513–518 (IEEE, 2015).

Tome, P., M. Vanoni and S. Marcel, “On the vulnerability of finger vein recognition to spoofing”, in “Biometrics Special Interest Group (BIOSIG), 2014 International Conference of the”, pp. 1–10 (IEEE, 2014).

Turing, A. M., “Computing machinery and intelligence”, Mind **59**, 236, 433–460 (1950).

Une, M., A. Otsuka and H. Imai, “Wolf attack probability: A new security measure in biometric authentication systems”, in “International Conference on Biometrics”, pp. 396–406 (Springer, 2007).

Vidyaratne, L. S. and K. M. Iftekharuddin, “Real-time epileptic seizure detection using eeg”, IEEE Transactions on Neural Systems and Rehabilitation Engineering **25**, 11, 2146–2156 (2017).

Wang, S.-Y., O. Wang, R. Zhang, A. Owens and A. A. Efros, “Cnn-generated images are surprisingly easy to spot... for now”, in “Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition”, pp. 8695–8704 (2020a).

Wang, T.-C., M.-Y. Liu, J.-Y. Zhu, G. Liu, A. Tao, J. Kautz and B. Catanzaro, “Video-to-video synthesis”, in “Advances in Neural Information Processing Systems (NeurIPS)”, (2018).

Wang, T.-C., A. Mallya and M.-Y. Liu, “One-shot free-view neural talking-head synthesis for video conferencing”, arXiv preprint arXiv:2011.15126 (2020b).

Wang, Y. and Y. Wang, “Cognitive informatics models of the brain”, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) **36**, 2, 203–207 (2006).

Wayman, J., A. Jain, D. Maltoni and D. Maio, “An introduction to biometric authentication systems”, in “Biometric Systems”, pp. 1–20 (Springer, 2005).

Wolpaw, J. R., N. Birbaumer, D. J. McFarland, G. Pfurtscheller and T. M. Vaughan, “Brain–computer interfaces for communication and control”, Clinical neurophysiology **113**, 6, 767–791 (2002).

Wu, Z., N. Evans, T. Kinnunen, J. Yamagishi, F. Alegre and H. Li, “Spoofing and countermeasures for speaker verification: A survey”, speech communication **66**, 130–153 (2015a).

Wu, Z., T. Kinnunen, E. S. Chng, H. Li and E. Ambikairajah, “A study on spoofing attack in state-of-the-art speaker verification: the telephone speech case”, in “Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference”, pp. 1–5 (IEEE, 2012).

Wu, Z., T. Kinnunen, N. Evans, J. Yamagishi, C. Hanilçi, M. Sahidullah and A. Sizov, “Asvspoof 2015: the first automatic speaker verification spoofing and countermeasures challenge”, in “Sixteenth Annual Conference of the International Speech Communication Association”, (2015b).

Xu, Y., T. Price, J.-M. Frahm and F. Monrose, “Virtual U: Defeating face liveness detection by building virtual models from your public photos”, in “25th USENIX Security Symposium (USENIX Security 16)”, pp. 497–512 (USENIX Association, 2016).

Yambay, D., B. Becker, N. Kohli, D. Yadav, A. Czajka, K. W. Bowyer, S. Schuckers, R. Singh, M. Vatsa, A. Noore *et al.*, "Livdet iris 2017—iris liveness detection competition 2017", in "2017 IEEE International Joint Conference on Biometrics (IJCB)", pp. 733–741 (IEEE, 2017a).

Yambay, D., J. S. Doyle, K. W. Bowyer, A. Czajka and S. Schuckers, "Livdet-iris 2013 - iris liveness detection competition 2013", in "IEEE International Joint Conference on Biometrics", pp. 1–8 (2014).

Yambay, D., L. Ghiani, P. Denti, G. L. Marcialis, F. Roli and S. Schuckers, "Livdet 2011—fingerprint liveness detection competition 2011", in "2012 5th IAPR international conference on biometrics (ICB)", pp. 208–215 (IEEE, 2012).

Yambay, D., B. Walczak, S. Schuckers and A. Czajka, "Livdet-iris 2015 - iris liveness detection competition 2015", in "2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)", pp. 1–6 (2017b).

Yampolskiy, M., P. Horvath, X. D. Koutsoukos, Y. Xue and J. Sztipanovits, "Taxonomy for description of cross-domain attacks on cps", in "Proceedings of the 2nd ACM international conference on High confidence networked systems", pp. 135–142 (ACM, 2013).

Yang, X., Y. Li and S. Lyu, "Exposing deep fakes using inconsistent head poses", in "ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)", pp. 8261–8265 (IEEE, 2019).

Zao, J. K. and et. al, "Pervasive brain monitoring and data sharing based on multitier distributed computing and linked data technology", Frontiers in human neuroscience **8** (2014).

Zeng, Y., Q. Wu, K. Yang, L. Tong, B. Yan, J. Shu and D. Yao, "Eeg-based identity authentication framework using face rapid serial visual presentation with optimized channels", Sensors **19**, 1, 6 (2019).

Zhang, R., N. Zhang, C. Du, W. Lou, Y. T. Hou and Y. Kawamoto, "From electromyogram to password: exploring the privacy impact of wearables in augmented reality", ACM Transactions on Intelligent Systems and Technology (TIST) **9**, 1, 1–20 (2017).

Zhang, X., S. Karaman and S.-F. Chang, "Detecting and simulating artifacts in gan fake images", in "2019 IEEE International Workshop on Information Forensics and Security (WIFS)", pp. 1–6 (IEEE, 2019).

Zhang, Y., J. Tian, X. Chen, X. Yang and P. Shi, "Fake finger detection based on thin-plate spline distortion model", in "International Conference on Biometrics", pp. 742–749 (Springer, 2007).

Zhang, Z., J. Yan, S. Liu, Z. Lei, D. Yi and S. Z. Li, "A face antispoofing database with diverse attacks", in "2012 5th IAPR international conference on Biometrics (ICB)", pp. 26–31 (IEEE, 2012).

Zhao, H., Y. Wang, Z. Liu, W. Pei and H. Chen, “Individual identification based on code modulated visual evoked potentials”, IEEE Transactions on Information Forensics and Security (2019).

Ziegler, A., E. Christiansen, D. Kriegman and S. J. Belongie, “Locally uniform comparison image descriptor”, in “Advances in Neural Information Processing Systems”, pp. 1–9 (2012).

Zúquete, A., B. Quintela and J. P. S. Cunha, “Biometric authentication with electroencephalograms: Evaluation of its suitability using visual evoked potentials”, in “Biomedical Engineering Systems and Technologies”, pp. 290–306 (Springer, 2011).