DriverAuth: A Risk-based Multi-modal Biometric-based Driver Authentication Scheme for Ride-sharing Platforms

Sandeep Gupta[a], Attaullah Buriro[a,c], Bruno Crispo[a,b]

[a]*Department of Information Engineering & Computer Science (DISI), University of Trento, Italy*
[b]*Department of Computer Science (DISI), DistriNET, KULeuven, Belgium*
[c]*Department of Information Security, KFUEIT Rahim Yar Khan, Pakistan*

Abstract

On-demand ride and ride-sharing services have revolutionized the point-to-point transportation market and they are rapidly gaining acceptance among customers worldwide. Alone, Uber and Lyft are providing over 11 million rides per day [1, 2]. These services are provided using a client-server infrastructure. The client is a smartphone-based application used for: i) registering riders and drivers, ii) connecting drivers with riders, iii) car-sharing to share the expenses, minimize traffic congestion and saving traveling time, iv) allowing customers to book their rides. The server typically, run by multi-national companies such as Uber, Ola, Lyft, BlaBlaCar, manages drivers and customers registrations, allocates ride-assignments, sets tariffs, guarantees payments, ensures safety and security of riders, etc. However, the reliability of drivers have emerged as a critical problem, and as a consequence, issues related to riders safety and security have started surfacing. The lack of robust driver verification mechanisms has opened a room to an increasing number of misconducts (i.e., drivers subcontracting ride-assignments to an unauthorized person, registered drivers sharing their registration with other people whose eligibility to drive is not justified, etc.) [3, 4, 5].

This paper proposes DriverAuth - a novel risk-based multi-modal biometric-based authentication solution, to make the on-demand ride and ride-sharing services safer and more secure for riders. DriverAuth utilizes three biometric modalities, i.e., swipe, *text-independent* voice, and face, in a multi-modal fashion to verify the identity of registered drivers. We evaluated DriverAuth on a dataset of $10,320$ samples collected from 86 users and achieved a True Acceptance Rate (TAR) of 96.48% at False Acceptance Rate (FAR) of 0.02% using Ensemble Bagged Tree (EBT) classifier. Furthermore, the architecture used to design DriverAuth enables easy integration with most of the existing on-demand ride and ride-sharing systems.

Keywords: Smartphone, Sensors, User Authentication, Physiological & Behavioral Biometrics, Risk-based Approach

1. Introduction

On-demand ride and ride-sharing services can deliver one-time rides to customers on a very short notice and are available 24×7 in all major cities worldwide. A customer can book a ride,

Email addresses: sandeep.gupta@unitn.it (Sandeep Gupta), attaullah.buriro@unitn.it (Attaullah Buriro), bruno.crispo@unitn.it (Bruno Crispo)

easily and quickly, through the dedicated smartphone-based ride-offering applications provided by different companies and downloadable at popular application stores, e.g., Play-store, App-store, etc. These services have facilitated quick business opportunities, allowing individuals to become partners (drivers) to offer rides to customers. However, the safety and security of the riders are always at risk, according to the news related to fake drivers and assaults by dishonest drivers [6]. On-demand rides and ride-sharing services, being easy to access and lucrative, are attracting people also with unclean police records to become driver-partners, by using false identities [7]. Currently, ride-sharing companies rely on the government-issued documents, e.g., passports, driver license, etc., to verify their drivers-partners identity and their eligibility to drive. However, this verification is generally performed only once at the time of their registration. Further, these documents are not difficult to forge [8] and not all countries use the same security standards to protect them. Most ride-sharing services support drivers' rating services on social-media such as Facebook, LinkedIn, Twitter, and Google+. However, these ratings can be easily manipulated, thus, not always reliable [9, 10, 11].

All these factors have contributed to an increasing number of incidents involving on-demand and shared rides in recent years [6, 12]. This trend has motivated ride-sharing companies to implement more rigorous checks on their drivers [13]. The checks that have been implemented, however, did not stop the abuses, e.g., dishonest drivers creating multiple accounts with forged documents [14]. These abuses are becoming also a liability concern [15], thus, the search for new, secure, and robust driver verification mechanisms becomes extremely important.

In spite of background checks on the drivers at the time of registration, the system lacks a robust mechanism [16], to verify the driver's identity each time she is offering a ride [5]. Some companies have introduced a real-time identity check that requires drivers to take a selfie before going online to drive [17] but not before each ride.

These open issues motivate the design of a new risk-based verification mechanism that can verify a legitimate driver at the time of every new registration and ride-booking, and thus, minimize the associated risks of abuses. An important requirement that any new driver authentication scheme must satisfy is not to alter the existing work-flow to pose a usability burden to drivers.

DriverAuth authenticates drivers by leveraging three biometric modalities, i.e., swipe, *text-independent* voice, and face, for verification purposes in a multi-modal fashion. Multi-modal systems are expected to be more reliable and accurate than unimodal systems, to verify a user. Furthermore, studies [18, 19, 20] have shown that multi-modal systems are more resilient to common attacks, e.g., presentation-, mimic-, replay-, random-attacks in comparison to unimodal systems.

The main contributions of this paper are as below:

- The proposal of DriverAuth- a multi-modal system that pro-actively verifies the drivers' identity every time drivers accept a new ride-booking. The proposed mechanism collects three biometric modalities, e.g., swipe gestures, *text-independent* voice and face, while they interact with the dedicated driver-application, to verify the drivers' identity. DriverAuth that can minimize the threat(s) posed by fake and malicious drivers. Hence, provisioning the safety and security of riders.
- Collection and sharing of swipe and voice data of 86 participants, for future research.
- Experimental evaluation of DriverAuth on the dataset of 86 users.

Paper Organization

The rest of the paper is organized as follows: Section 2 covers the related work. Section 3 describes the problems in the existing driver registration process and the risk involved in this system along with the need of risk-based user verification method and the considered threat model. Section 4 presents DriverAuth design including the verification process at the time of new registration and ride-booking assignment. Section 5 discusses the methodology used to collect the dataset, to extract features, to concatenate and selection of the best features from the chosen modalities. Section 6 covers the details of the experiments, the classification method, and presents the performance evaluation and the obtained results. Finally, Section 7 concludes the paper outlining possible future work.

2. Related Work

Face recognition is one of the most widely accepted biometric modality mainly because it provides high recognition rates. Thus, Uber has introduced "Real-Time ID Check" - a face recognition system developed by Microsoft, to verify the identity of their registered drivers [17]. The system collects the face images of the person registering as driver-partner, extracts facial features, and store them in the database for future verification purposes. Only a subset of randomly-selected driver-partners are asked to verify themselves using "Real-Time ID Check". Selected drivers are requested to take a selfie, then, this query image is compared with reference images to verify their identity. Subsequently, the system takes necessary action, i.e., allows/disallows drivers to offer rides, based on the obtained verification results from the face recognition algorithm. Uber claims 99% success rate of this mechanism, however, they have not yet published any details related to their systems' robustness against presentation attacks and about liveness detection.

Multi-modal biometric factors can remarkably improve identity verification accuracy of a system by combining the pieces of evidence extracted from single modalities [32]. Multi-modal systems are also more resilient against spoofing in comparison to unimodal ones [33]. Our system is the first multi-modal biometric authentication scheme to address driver's authentication problem for ride-sharing services. Similar proposals exist but only for user authentication on smartphones. Table 1 summarizes the most relevant multi-modal user authentication solutions on smartphone.

Proteus, proposed by Gofman et. al. [19], is a bi-modal biometric verification system based on face and voice features, for mobile devices. This scheme extracts principle components using Principal Component Analysis (PCA) and Mel Frequency Cepstral Coefficients (MFCC) from face and voice modality, respectively, to construct a bi-model system. The system was evaluated on a dataset of 54 users and it achieved an Equal Error Rate (EER) of 2.14% using latent Dirichlet allocation (LDA) fusion method. Another bi-modal approach [21] incorporates finite Gaussian Mixture Model (GMM) based on Expectation Maximization (EM) and applies score-level fusion to fuse face and voice modalities. They achieved an EER of 0.449% for face and 0.003% for voice modalities, in unimodal settings, and their bi-modal settings yielded an EER of 0.087%, on the dataset of 30 participants. These experiments clearly reflect the potential of multi-modal biometrics to enhance the verification accuracy on mobile devices.

Swiping is a very common gesture required to interact with mobile devices' touchscreen. It is a collection of touch-points generated while the user dragged her finger on the smartphone touchscreen [34, 35, 36]. Feng et al. [25] proposed Finger gesture Authentication System using Touchscreen (FAST). They applied Random Forest as classifier and achieved a FAR of 4.66% and

Table 1: Multi-modal (combination of face, voice, or touch) User Authentication Schemes

Reference	Modalities Used	Algorithms Used	Dataset Size	Performance
Gofman et al. [19]	Face, Voice	Latent Dirichlet allocation (LDA) fusion method	54	EER=2.14%
Soltane et al. [21]	Face, Voice	Finite Gaussian Mixture Model (GMM) based on Expectation Maximization (EM) using score-level fusion	30	EER=0.087%
Wang et al. [22]	Face, Voice	Quantization Index Modulation (QIM) and Gaussian Mixture Models (GMM)	295	EER=2.76 – 3.79%
Menzai et al. [23]	Face, Voice	Dempster-Shafer theorem using belief function	295	HTER=0.433 – 2.875%
Kim et al. [24]	Face, Voice	Generalized cross correlation (GCC) algorithm and AdaBoost algorithm on Local binary pattern	-	Accuracy=95%
Menzai et al. [25]	Face, voice	Belief functions and Particle Swarm Optimization (PSO)	295, 52	EER=0.5 to 0.9
Feng et al. [25]	Finger gesture Authentication System (FAST)	Random Forest classifier	40	FAR=4.66%, FRR=0.13%
Buriro et al. [26]	Swipe, Pickup movement, and Voice	Bayesian classifier	26	HTER=7.57%
Aronowitz et al. [27]	Fingertip-based writing, Face and Voice	Dynamic time warping (DTW)	32	EER=0.1% at quiet place, and 0.5% in noisy surroundings
Akhtar et al. [28]	Face, touch-stroke, and the hands-movements to holding phone	Multilayer Perceptron (MLP)	95	EER=1%
Buriro et al. [29]	Touch-tapping and hands-movements to holding phone	Multilayer Perceptron (MLP)	97	TAR=85.77
Koreman et al [20]	Voice, face and signature	Gaussian mixture models (GMMs)	82	EER=2%
Buriro et al. [30]	Touch-typing and hands-movements to holding phone	Multilayer Perceptron (MLP)	95	TAR=96
Eastwood et al. [31]	Face, iris, and fingerprints	Belief (Bayesian) networks	-	-

False Reject Rate (FRR) of 0.13% for the continuous post-login user authentication on a dataset of 40 users. ITSME [26] - a multi-modal authentication mechanism utilizes three behavioral modalities (swipe, pickup movement, and voice) and by applying Bayesian classifier achieved 7.57% Half Total Error Rate (HTER) on their collected dataset of 26 participants. Another proposal by Aronowitz et al. [27], combines user's fingertip-based writing on multi-touch screens with face and voice features and uses dynamic time warping (DTW) engine for user verification. They achieved an EER of 0.1% at quiet place, and 0.5% in noisy surroundings, on their collected dataset of 32 users (20 males and 12 females).

Akhtar et al. [28] leveraged face, touch-stroke, and the phone-movements (the phone's micro-movements generated while the user types her secret), to propose a multi-modal user authentication solution for smartphones. It is worth noting that authors collected touch-stroke, and the corresponding phone-movements data by themselves and relied on MoBio dataset [37], for face modality to generate a tri-modal chimerical dataset. The experiment was conducted on 95 subjects and yielded an overall EER between 1% to 4% for a trimodal system using Multilayer

Perceptron (MLP) and Random Forest (RF) as classifiers. Another similar effort [20] leverages voice, face and signature modalities, for user authentication on mobile devices. This approach yielded an EER of 2% using Gaussian mixture models (GMMs). The system utilized BANCA audio-visual database [38] and BIOMET on-line signature database [39] comprising of the data collected from 82 and 84 subjects, respectively. Authors also checked each modality in unimodal settings and achieved an EER of 28%, 5%, and 8%, for face, voice, and signature modalities, respectively. The fusion of three modalities enhanced the system accuracy and reduced the EER to just 2%.

Liveness detection is generally deployed to detect spoofing attacks. According to Zhang et al. [40], mobile audio hardware can be used to exploit articulatory gesture of a user to detect liveness and their proposed "VoiceGesture" system achieves 99% detection accuracy at approximately 1% EER. Swipe gesture is the result of a user subconscious muscle memory involving a sweeping movement on the touchscreen developed over a period of time due to constant use of a smartphone. Swipe gestures are arguably considered hard to be imitated and the impostor's attempts are easily detectable [41]. Also, swipes have no explicit visual indicators which make it furthermore resistant to mimicry attacks [42]. Lastly, it is comparatively easier to perform liveness detection on faces because some of the robust liveness detection methods are already available [43], to prevent face spoofing attacks [44].

Our proposed scheme DriverAuth is different from existing state-of-the-art in several ways: firstly, DriverAuth is a client-server-based multiuser (`multiclass`) verification solution in contrast to the existing multimodal systems [19, 28, 26]. More specifically, we model this as a multi-class classification problem (classifier training with multiple users) whereas, the existing approaches dealing with smartphone user authentication are one-class or binary class classification problems. Secondly, DriverAuth utilizes both physiological and behavioral biometric modalities, i.e., swipe, face, and *text-independent* voice, equipped with liveness detection as a result more resilience to spoofing.

3. Problem Description

On-demand ride and ride-sharing services have revolutionized the point-to-point transportation market, in a short period of time. Technology-based companies, e.g., Uber, Ola, Lyft, Blablacar, Sidecar, etc., connect customers and drivers by means of dedicated smartphone-based applications. Customers interested to the services and individuals aspiring to become driver-partner, can download these dedicated applications free-of-cost, available at online-app-stores, e.g., Play-store, App-store, Microsoft-store, etc.

In order to become a driver-partner, an individual needs to be older than 21 years old, should be in possession of the valid driving license, valid vehicle registration, clean driving record, and have no criminal history [45]. These background checks are performed by the service provider just once, prior to the registration. Once the individuals are accepted as driver-partners, they can accept rides' requests, reserved by customers, using dedicated driver-application on their smartphones and perform their duty. Surprisingly, the system providers do not verify their drivers' identity while they accept a new ride, requested by the customers [46]. Thus, system providers are neither able to monitor fake drivers [3] nor they are able to curb dishonest drivers with multiple identities [47]. Therefore, the safety and security of the customers are always at risk and this risk in increasing with the increasing number of abuses reported every year [6].

The safety and security of a customer is a huge challenge in on-demand ride and ride-sharing systems, despite being convenient, fast, and economical. Considering the volume of rides (alone, Uber and Lyft are providing over 11 million rides per day [1, 2]), even if only one rider in a million is victimized, this sum up to 11 victims per day. As driver-partners can join and leave the service at anytime without any obligation is difficult to deter abuses.

3.1. Threat model

We consider two different types of malicious users in our scenario: the first type of adversary can impersonate a driver-partner by imitating a legitimate driver. The second type of attacker colludes with a legitimate driver-partner and share with him/her the registration to provide rides on behalf of the legitimate driver.

Both adversarial situations can be countered using DriverAuth. DriverAuth leverages swiping, voice and face combined together to verify the legitimate driver at run-time and would require driver's presence every time she accepts a new ride request. Additionally, the fusion of the three modalities increases the resilience to common attacks, i.e., presentation, mimic and replay attacks [18, 19, 20].

3.2. Risk-based verification mechanism

According to ISO 9000:2015 [48], *risk* is the "effect of uncertainty on objectives". The *objectives* can be defined as the strategic, tactical, or operational requirements pertaining to an ecosystem. Whereas, the *effect* can cause both positive or negative deviations on the objectives. A *risk-based verification mechanism* aims at determining uncertainties to minimize their effects on the set objectives.

At present, on-demand ride and ride-sharing services use the concept of simple verification mechanism [49], in which, users are verified at the time of entry only, and users are considered legitimate until they quit the system. However, with reference to the threat model, discussed in section 3.1, the drivers' verification at the time of each new ride-assignment becomes imperative, to ensure customers safety and security. In that case, a simple verification concept does not suffice owing to their limitations to prevent potential risk hazards. Therefore, a risk-based verification mechanism could be the potential solution.

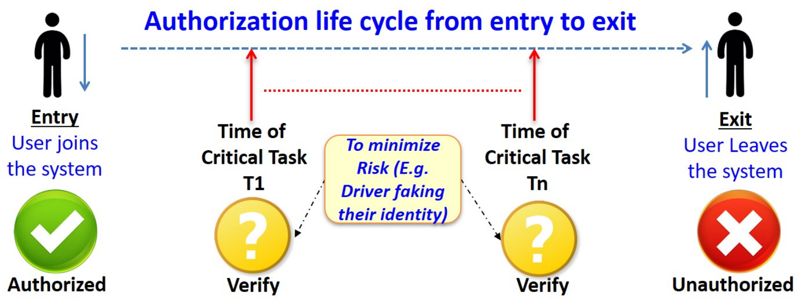


Figure 1: Risk-based verification mechanism

The life cycle of a typical risk-based verification mechanism consists of users authorization at the time of entry and their verification at every critical operation. As illustrated in Figure 1, users can be authorized to use the system by registering to it, i.e., *Entry*, and once they unregistered themselves, i.e., *Exit*, they are unauthorized to use the system. At the time of registration, users are added to the database for a reliable $1 - to - 1$ verification. Every time $(T_1...T_n)$ users carry out a critical activity (e.g., accept a ride request) they are verified regardless of the fact that they are legitimate drivers. If an incident is reported, it is added to the incidents database tagged with the responsible user identity, for future reference.

The concept used in Risk Profiling tools [50, 51] to assess risk at different stages of a critical system can be applied here for proactive risk assessment [31] by analyzing the incidents database. This incidents database can be further utilized for Evidence Accumulation and Risk Assessment (EA&RA) to evaluate the driver's behavior in the past and present using special risks indicators [52]. However, we consider risk assessment as our future work.

4. Our Solution: DriverAuth

DriverAuth authenticates the drivers at the time of registration and at the time of new ride-assignments. Each service provider has their own dedicated system and application for their driver-partners, however, the core functionalities are the same. Thus, DriverAuth can easily be integrated into these systems and provide the required safety and security to customers.

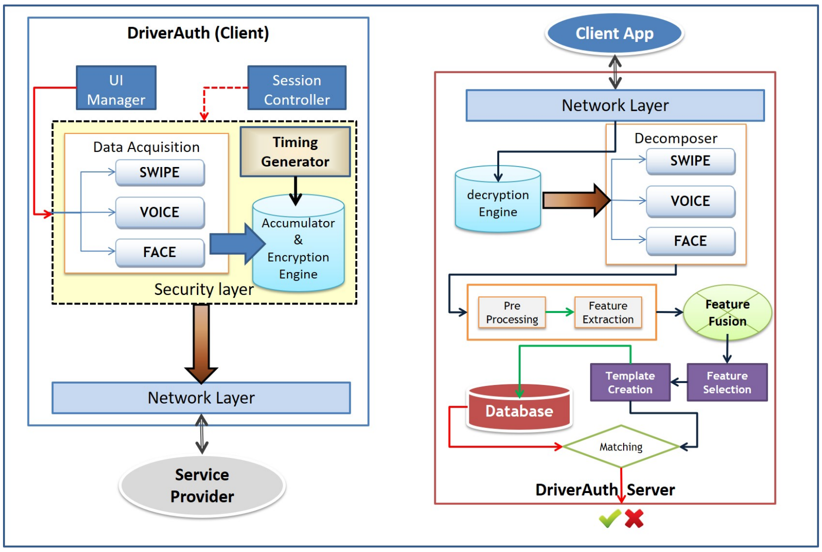


Figure 2: DriverAuth architecture [53]

DriverAuth uses the client-server architecture [53] as illustrated in Figure 2. The client application consists of a data acquisition module, accumulator/encryption engine, and a timing generator. Data acquisition module collects the swipe data, voice-print, and face-image in sequential manner using blocking-call-mechanism, i.e., application allows to proceed only after it receives the required user's input. The operational details of the data collection process for driver verification is described in Section 4.3. The data collected, i.e., touch-points data, $2 - seconds$ voice-prints, and a face image, are temporarily stored by accumulator and encryption engine module for encryption, packaging, and time-stamping. With no delay, data is transferred to the server.

The server side consists of a) a decryption engine, b) a decomposer, c) signal preprocessing, d) features extraction module, e) feature fusion module, f) feature selection module, g) template creation module, and h) database module. Decryption engine decrypts the user-data as received from the client application, which is further decomposed into individual modalities. As the proposed scheme uses the multi-modal mechanism, features are fused and selected on merit basis entailing the selection of only productive features for user authentication. The drivers template is created based on the selected features subset and is then stored in the central database as training templates with a proper label. Later, a similar procedure is applied to the testing data to generate the testing template. In order to verify the identity of the claimant, the testing template is matched against the existing labeled training templates, present in the database.

4.1. **DriverAuth** ***Design***

On-demand ride and ride-sharing systems have three primary stakeholders: a) centralized smartphone-based administration , b) customers and c) drivers, as illustrated in Figure 3.

DriverAuth verifies the person both at the time of registration and at new ride-assignments. A security layer is stitched to the driver application to collect the biometric modalities, e.g., voice, swipe gesture, and face. Simultaneously, the captured data (query input) is transferred to the server for driver's identity verification. Also, this query input can be looked up in the stored database for any incident flagged against it.

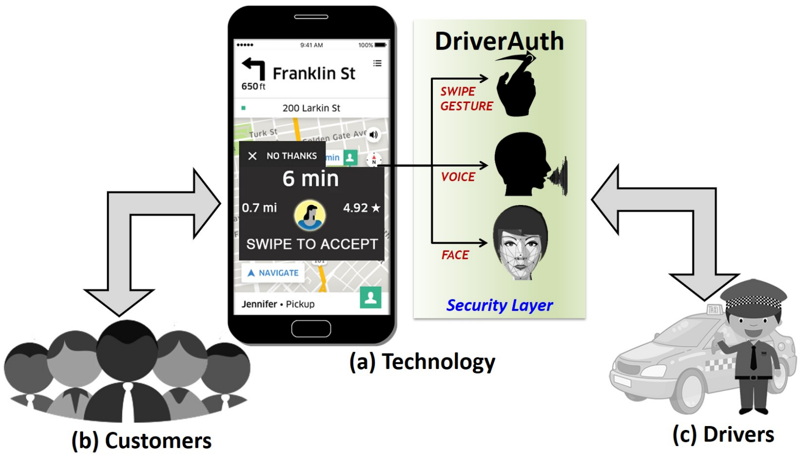


Figure 3: On-demand ride and ride-sharing system stakeholders

4.2. ***Verification during driver-partners registration***

Verification process during driver-partners registration is illustrated in Figure 4.

1. Individuals can apply to become driver by filling the application form using dedicated driver-application (see Figure 4) on their smartphone.

2. During the registration process, DriverAuth collects the swipe gesture, *text-independent* voice and face samples of a person.

3. At the server (see Figure 4), query input is first compared with the stored driver-partner templates in the database. If this query input is positively verified, the registration is completed. If there is a new registration, the new template is added to the database confirming the new registration.

Thus, DriverAuth minimizes the threats posed by dishonest drivers by preventing multiple or forged account creation.

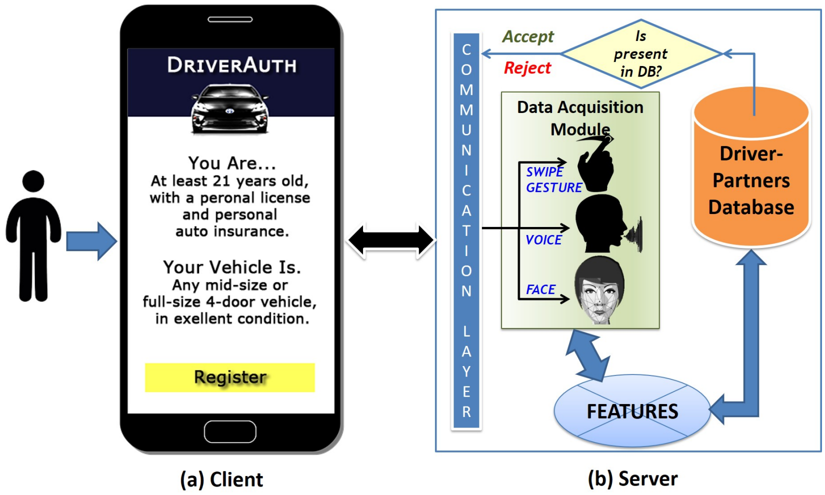


Figure 4: Overview of driver-partners registration process

4.3. *Verification during new ride assignment*

Drivers verification process during new ride booking is illustrated in Figure 5.

1. The customers can book the ride by setting up their location using the dedicated on-demand ride and ride-sharing application on their smartphones. Subsequently, they can locate the available cabs (along with driver's picture and vehicle details) near to their location to reserve the ride by selecting one of the cab [54].

2. On receiving a booking request from a customer, system provider forwards the request to the respective driver.

3. The driver upon receiving the alert can continue to accept the new ride-assignment by swiping on the touchscreen.

4. After the swipe input is detected, the application requires a short voice-print ($2-seconds$ of voice recording) from the driver. This voice-print can be totally *text-independent* that provides flexibility to the drivers to use any language of their choice.

5. After the successful voice detection, the application turns on the camera and prompts for the driver's selfie to conclude the ride-assignment acceptance process.

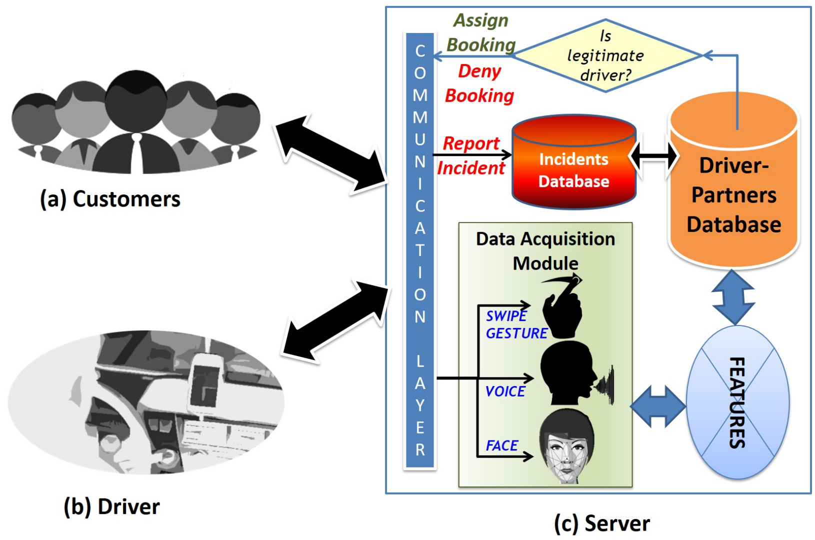


Figure 5: Overview of new ride assignment process

6. Subsequently, DriverAuth client application transfers the encrypted driver's biometric modalities, i.e., swipe gesture, voice, and face, to the server. In the meantime, the driver verification is performed on the server.

7. Based on the driver verification results, the system provider can approve the ride-assignment to the respective driver and simultaneously, intimate the customer.

8. In any case if driver abuses or assaults the rider, then the rider can report the incident, immediately. The reported incident can be tagged with the driver's identity which will automatically be added to the incidents database.

DriverAuth minimizes the potential risks towards the safety and security of riders by verifying the drivers' identity pro-actively, at the time of every new ride assignment.

4.4. Liveness detection and preventing spoofing attacks

Liveness detection helps to distinguish between living and non-living, during the authentication process and, thus, prevents spoofing attacks at the data acquisition module [55]. DriverAuth data acquisition module acquires data from three modalities, i.e., voice, swipe, and face, as described in Section 4.3. For voice liveness detection, data acquisition module incorporates phoneme sound localization mechanism taking advantage of the users unique vocal system and high quality stereo recording of smartphones [56]. Studies have shown that swipe gesture is inherently difficult to spoof [42] but in future we will incorporate technique for swipe liveness detection too. Similarly, face modality liveness indicators like eye blinking, mouth movements, face posture and motion analysis etc., are exploited for multi-spectral and reflectance analysis [57].

Thus, DriverAuth prevents the spoofing or presentation attacks at the sensor level by utilizing available mechanisms to detect liveness for each modality.

5. Methodology

DriverAuth exploits three biometric modalities, i.e., swipe gestures, voice, and face, and collects their corresponding data, while the users interact with a driver-application on their smartphones. Both physical and behavioral biometric modalities can be easily collected using smartphone's built-in hardware sensors, such as, camera, microphone, and touchscreen. We modeled this remote-user-verification as a *multi-class classification problem* because the scenario demands simultaneous classifier training and testing for multiple drivers, however, each query input needs to be assigned only to one class.

5.1. Datasets

We evaluated DriverAuth on a collected dataset of 86 recruited users. We developed a customized Android application to collect the swipe gestures and voice data. We outsourced the data collection activity to Ubertesters[1] - a crowd-sourcing platform to collect these two modalities in an unsupervised environment (in the wild) and they recruited more than 150 experienced professional testers worldwide for our experiment. However, some participants were rejected for several reasons: firstly, their smartphones were not found compatible with our experiment because they did not have the required sensors, secondly, they could not complete the experiment as instructed, and lastly, their data was noisy. For face data, we relied on MoBio database [37]. As all three modalities are mutually independent of each other, we augmented them to form a single dataset [58]. Thus, we created a chimerical dataset by associating these three modalities, i.e., swipe gesture, voice, and face to perform the analysis.

5.1.1. Swipe & Voice Data Collection

The prototype application was developed for Android OS (OS version `4.4.x` and above). It uses built-in hardware, i.e., touchscreen and microphone, to acquire touch points data during swipe action and recording of user's voice. We collected in total $10,320$ samples. The experiment was conducted in 4 sessions over the span of 3 days. Each user trained the application for 90 times in 3 sessions (30 times per session) within 15 minutes each. In fourth session, each user tested the application for 30 times. A total 120 observations were collected per user with $7,740$ (86×90) training samples and $2,580$ (86×30) testing samples.

As our developed application uses client-server architecture, the data generated as result of user's actions, i.e., swipe and voice command, is encrypted and zipped on the client device, i.e., smartphone, and is automatically transferred to the server, for further processing. On-demand ride and ride-sharing companies are operating worldwide.

Our prototype collects $2-$ *seconds text-independent* voice-print (e.g., "I accept the ride to Y"), allowing drivers to interact in the language depending on the country where they operate or the company for which they work. Therefore, we do not limit voice modality to any specific language or the particular word-sets.

Table 2 presents the demographics data of users participated in this experiment. Among 86 participants, 56 were males, 29 were females with 77 right-handed and 9 left-handed. Majority of participants were in Asia (28) and Europe (52) while performing the experiment, with 60 were between 20 to 30, 17 were between 30 to 40, and 3 were above 40.

[1] https://ubertesters.com

Table 2: User demographics

#	Parameter	Description
1	No. of Users	86
2	Gender	56 males, 29 females, 1 undisclosed
3	Handedness	77 Right, 09 Left
4	Age Groups	[20 to 30] - 60, [30 to 40] - 17, 40 plus - 3
5	Participants Location	Asia - 28, Europe - 52, North America - 5, South America - 1

5.1.2. MOBIO Dataset

This public dataset consists of face samples collected from 152 subjects in 2 phases using a NOKIA N93i mobile phone under realistic and uncontrolled environment over a period of 18 months from six sites across Europe [37]. In the first phase, 21 videos per participant were collected, whereas 11 videos per participant were acquired in the second phase. The data acquisition were spread over 6 different sessions per phase for each participant. The database has 1 : 2 female to male ratio, approximately. However, we picked only 86 subjects out of 150 to match the same number of users as to our dataset.

5.2. Feature Extraction

In this section, we explain the extraction of features for all the three selected modalities using statistical methods. Univariate statistical properties, i.e., mean, standard deviation, kurtosis or skewness has several benefits, they reduce the dimensionality of raw data, improve the signal-to-noise ratio, and they can be processed efficiently [59].

- **Swipe Modality:**
  A sequence of touch-events is generated every time user swipe on smartphone touchscreen using their finger. These touch-events are collected and encoded as an input sequence of finite length (n). Where, each sequence contains several attributes like time-stamp of the touch event (t_n), x-and y-coordinate of the touch point (x_n, y_n), pressure calculating how hard the finger was pressed on the screen (p_n), and size of touch area (s_n). We processed the collected sequences and extracted 33 features as listed in Table 3. The final feature vector is the concatenation of all the 33 features.

Table 3: List of swipe features

No.	Swipe Features			
1-4	Duration 1	Average event size 2	Event size down 3	Pressure down 4
5-8	Start X 5	Start Y 6	End X 7	End Y 8
9-12	Velocity X Min 9	Velocity X Max 10	Velocity X Average 11	Velocity X STD 12
13-16	Velocity X VAR 13	Velocity Y Min 14	Velocity Y Max 15	Velocity Y Average 16
17-20	Velocity Y STD 17	Velocity Y VAR 18	Acceleration X MIN 19	Acceleration X Max 20
21-24	Acceleration X AVG 21	Acceleration X STD 22	Acceleration X VAR 23	Acceleration Y MIN 24
25-28	Acceleration Y Max 25	Acceleration Y AVG 26	Acceleration Y STD 27	Acceleration Y VAR 28
29-32	Pressure Min 29	Pressure Max 30	Pressure AVG 31	Pressure STD 32
33	Pressure VAR 33	-	-	-

- **Voice Modality:**
  The voice signal contains 2 channels sampled at 44100 Hz with 16 bits per sample. The

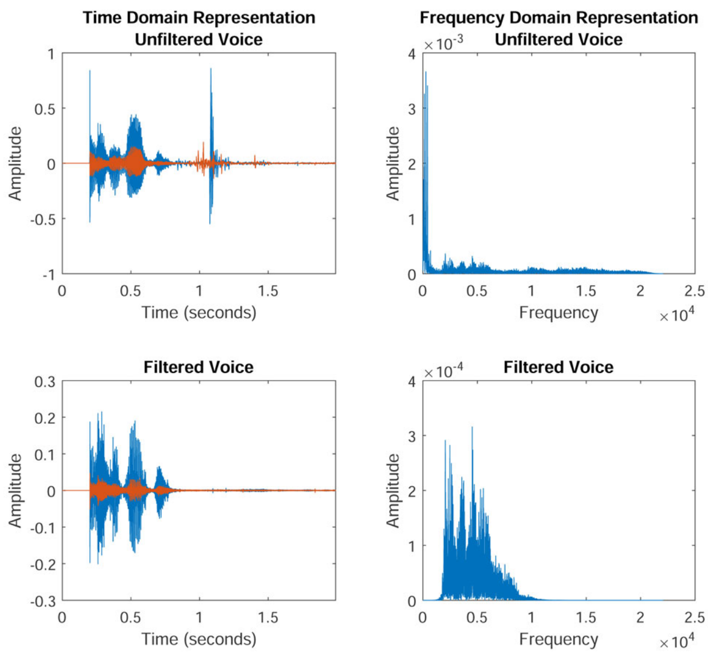


Figure 6: Voice signal filtering result

signal is first filtered using a bandpass filter. It can be observed in Figure 6 that by applying bandpass filter there is a significant improvement in signal-to-noise ratio.

Then, we computed MFCC [60] from these filtered voice signals. MFCCs are analogous to filters (vocal tract) in the source-filter model of speech. Relatively, the frequency response of vocal tract is smoother than the source of voiced speech. Thus, the vocal tract can be estimated by the spectral envelope of a speech segment. This technique is often used in voice recognition because it tracks the invariant feature of human speech among different persons.

Figure 7 illustrates the MFCCs computation process. After improving the signal-to-noise ratio, Fourier transform of a window of the voice signal is performed, then scaling of frequency axis to the non-linear Mel scale (using triangular overlapping windows) is done. In the next step, Discrete Cosine Transform (DCT) is performed on the log of the power spectrum of each Mel band. The MFCCs are the amplitudes of the resulting spectrum, which is a $2-D$ vector of size $13 \times \textit{variable length}$ (the length of vector depends on the voice signal duration).

We computed 4 statistical features, namely mean, standard deviation, kurtosis, and skewness, from a 2-D MFCC vector. Thus, the total 8 statistical features each of size 1×13 are generated from each left and the right voice channel. Finally, these 8 vectors of size 1×13 are concatenated to form a single $1-D$ feature vector of dimension 1×104.

- **Face Modality:**

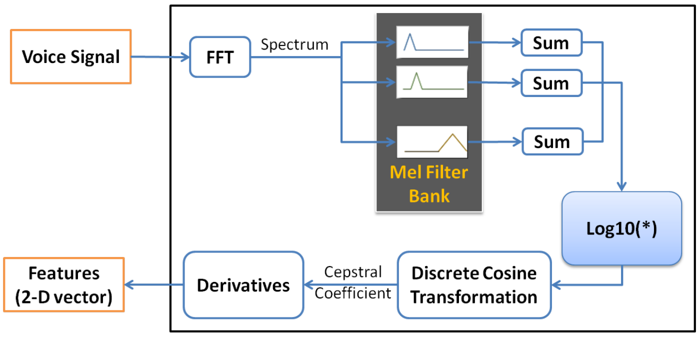


Figure 7: Voice features: MFCC computation process

On the server, the region of interest (ROI) is extracted automatically, by cropping the original images, as illustrated in Figure 8. Then, each image is converted into 8-bit grayscale format. We used the *Binarized Statistical Image Features* (BSIF) filter to obtain statistical features [61].

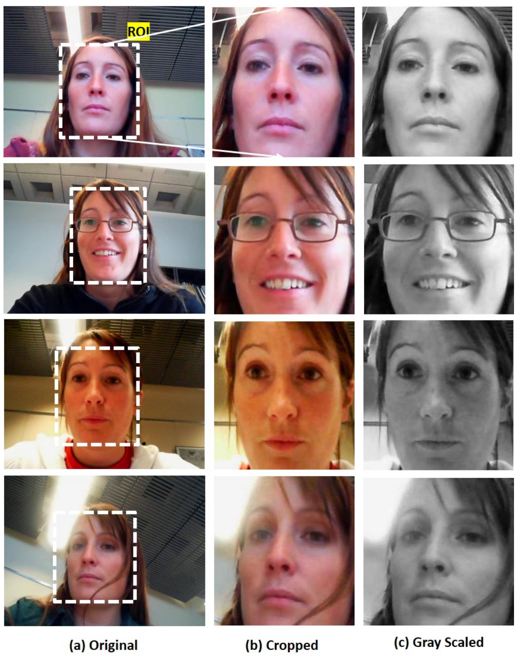


Figure 8: Face features: BSIF computation process

Given an image patch X of size $l \times l$ pixels and a filter W of size $n \times n$ pixels, where n is less than l. The filter response s_i can be obtained as shown in Equation 1.

$$s_i = X[l,l] * W[n,n] \tag{1}$$

We extracted 256 features per image using filter of size 3×3 with 8 bits word-length.

BSIF filter applies learning, instead of manual tuning, to compute statistically meaningful representation of an image.

5.3. Features Concatenation

Data fusion in a biometric system is a process of integrating multiple modalities to produce more accurate, consistent, and comprehensive information of users. Biometric researchers often consider that the early data fusion increases the accuracy of the system [62, 19]. However, sensor-level fusion does not yield the best results owing to the presence of noise during data acquisition. Thus, feature-level fusion is a better choice to improve the accuracy of the system, because feature representation reflects more relevant information on users. Lastly, this setting is preferred as it combines independent modalities [63]. Therefore, we applied feature level concatenation to generate the final features vector.

5.4. Feature Subset Selection

Feature selection plays an important role in fine-tuning of the chosen classifiers. It helps in reducing the dimension of data as well as prevent the over-fitting by identifying productive features out of the full feature-set. This process not only maximizes the accuracy of a classifier but also contributes to improving classifier's decision-making time. Feature selection methods can be categorized as *Filter*, *Wrapper*, *Embedded*, and *hybrid* methods, based on their relationship with the construction of a model [64]. We considered Information Gain Ranking Filter[65], Simple Correlation Ranking Filter [65], CFS Subset Evaluator with greedy forward search [65], and ReliefF [64] to obtain most productive feature subset, for our analysis. However, relief-based algorithms (RBAs) provided the best accuracy result.

RBAs belong to the individual evaluation $filter$ method. The advantages of RBAs are: 1) they are able to detect conditional dependencies between features, 2) they provide a unified view on the features estimation in classification, and 3) they are relatively faster (with an asymptotic time complexity of order $O\,(instances^2 \times features)$) to other feature selection methods [64, 66].

RBAs compute ranks and weights of features to derive feature statistics using the concept of nearest neighbors as shown in Equation 2.

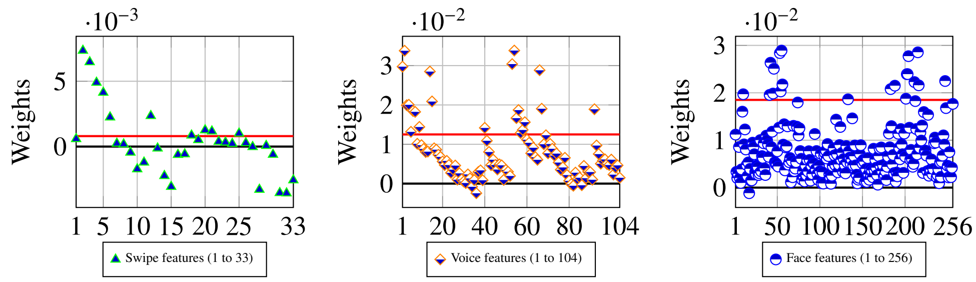
$$[RANKED, WEIGHT] = relief(X, Y, K) \tag{2}$$

Where, X $(m \times n)$ is a given 2-d dataset, Y $(m \times 1)$ is the response vector, and K is a number of nearest neighbors. $RANKED$ are indices of columns in X ordered by attribute importance, meaning $RANKED[1]$ is the index of the most important feature. $WEIGHT$ are features weights ranging from -1 *to* $+1$ with large positive weights assigned to most important attributes.

We performed feature selection in three settings and evaluated DriverAuth in unimodal, bimodal, and trimodal settings. Then, we tested and validated our system on both full feature set and selected feature set to achieve an optimal design. In the following sections, we explain our feature selection strategy for our experiments.

- **Unimodal:** We obtained in total 33, 104, and 256 features from processed swipe, voice, and face modalities, respectively, to design the unimodal systems. We evaluated the system firstly on the full feature set. To evaluate the system on the selected feature set, we estimated the importance of features of each modalities using ReliefF algorithm[2]. Then,

[2] https://in.mathworks.com/help/stats/relieff.html

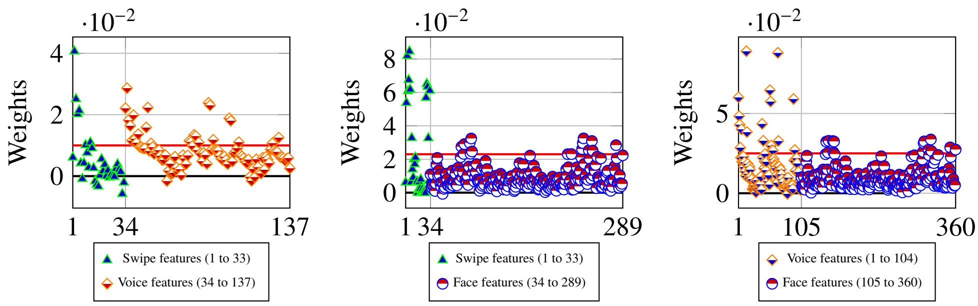


(a) Swipe features (11) (b) Voice features (20) (c) Face features (20)

Figure 9: Unimodal System: Plot between features vs. weights

we picked top 30% or 20 features of the total (whichever is less) as per their weights. The features vs. weight for the three modalities are shown in Figure 9.

The number of features required for the best classification model creation was computed, empirically. In case of swipe, the total number of features available are 33, we, firstly, trained our classification model by picking all the features with positive rank, i.e., above zero as shown in Figure 9a and observed that the same TAR is achieved with top 11 features, i.e. 33% of total available features as demarcated by a red line in Figure 9a. Whereas, in case of voice and face, the classification model is trained by picking top 33% of total available features, i.e., 34 and 85 features, respectively. But, we observed that with only top 20 features the same TAR is achieved as demarcated by a red line in Figure 9b and Figure 9c.



(a) Swipe + Voice features (31) (b) Swipe + Face features (40) (c) Voice + Face features (40)

Figure 10: Bimodal System: Plot between features vs. weights

- **Bimodal:** We concatenated swipe and voice, swipe and face, and voice and face creating feature set of dimension 137, 289, and 360, respectively, to design a bimodal system. In this case, for each combination, the two feature sets are firstly fused and then ranked using ReliefF algorithm. Finally, the system is evaluated on full and selected feature set. The dimension of selected features for swipe + voice, swipe + face, and voice + face are 31, 40, and 51, as demarcated by a red line in Figure 10a, Figure 10b and Figure 10c, respectively.

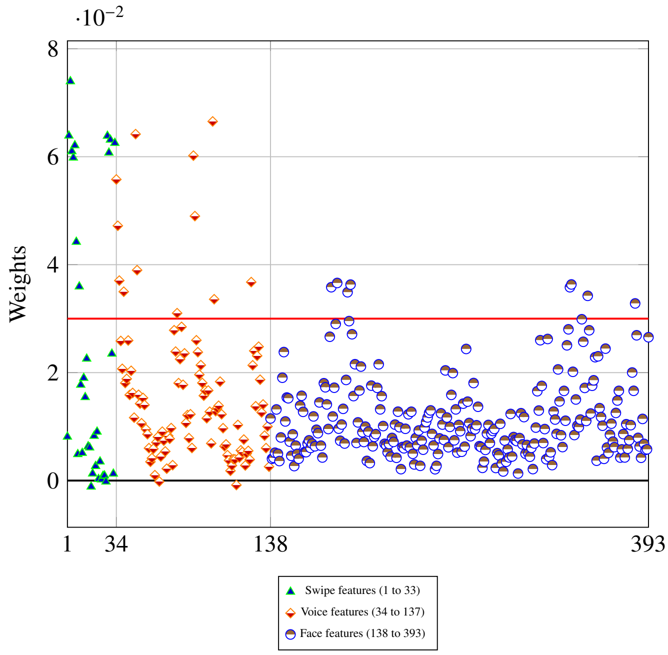


Figure 11: Trimodal system: Plot between features vs. weights (51)

- **Trimodal:** We concatenated the feature sets of each modality together to create a single feature set of dimension 393 for evaluation of DriverAuth in trimodal settings. Figure 11 represents features ranking obtained by applying the ReliefF algorithm on the fused feature set. Finally, the system is evaluated on both full and selected feature set of dimension 51.

6. Validation

We utilized Classification Learner [67] to generate a classification model. Classification Learner can perform automated training to search for the best classification model type, e.g., support vector machines, nearest neighbors, ensemble classification, etc. We used 5-fold cross-validation to assess the predictive performance. Cross-validation protects against over-fitting by partitioning the data set into folds and estimate accuracy on each fold. Thus, this method gives the good estimation of the predictive accuracy of the final model trained with full data.

However, security-sensitive infrastructures, e.g., banks, prefer to design classification models with fewer number of training samples (typically up to 10). Thus, we evaluated our trimodal system with most productive feature-set achieved by applying the `ReliefF` algorithm for a different number of training samples, i.e., 10, 20, 30, and 40, to determine its effectiveness. To achieve it, we split the dataset into two parts, i.e., training- and testing- datasets and evaluated the model in two different scenarios. In the first scenario, we utilized a designated number of training samples (n) to train the classifier and used $120 - n$ samples to test the model. Here, we presented the result in terms of `TAR`, which can be further studied in Figure 12. In the second scenario, i.e., the zero-effort attack scenario (where an impostor could only make random tries to access the system without knowing the actual user), we excluded legitimate samples, i.e., 120 samples, of each user and used the remaining samples, i.e., 10200 (85 × 120) to attack the model, for all the remaining 85 users. Here, we presented the results in terms of FAR, which can be further studied in Figure 13.

6.1. *Classification Methods*

In a biometric system, the role of a classifier is to recognize the similarities, or detect the anomalies between the query input and stored templates to authenticate a user. We selected Support Vector Machines, Nearest Neighbor, and Ensemble classifiers to evaluate DriverAuth, using multi-class classification model. These classifiers are well suited for the multi-class environment and have shown to be very effective for similar biometric modalities, i.e., swipe, voice, and face, in recent studies [26, 25, 61, 63].

Table 4: Classifiers comparison.

Classifier Type	Algorithm	Prediction Speed	Memory Usage
Quadratic SVM	Finds the best hyperplane that separates data points of one class from those of the other class	Slow	Large
Ensemble Bagged Trees	Random forest Bag, with Decision Tree learners	Medium	High
Weighted KNN	Medium distinctions between classes, using a distance weight.	Medium	Medium

Table 4 lists our chosen classifiers and compares them in term of their prediction speed and memory usage (for more details on the classifier benchmarking refer to [68]).

6.2. *Performance Evaluation*

We use the following metric to report our results:

- **True Acceptance Rate (TAR):** It is a ratio of correctly accepted owner's attempts to all the attempts made [69]. Higher TAR indicates that the system performs better in recognizing a legitimate user.
- **False Rejection Rate (FRR):** It is a ratio of incorrectly rejected attempts of a legitimate user to all the attempts made [69]. It is calculated as `FRR = 1 - TAR`.
- **False Acceptance Rate (FAR):** It is a ratio of incorrectly accepted impostor attempts to all the attempts made [69]. Lower FAR means the system is robust to impostor attempts.
- **True Rejection Rate (TRR):** The ratio of correctly rejected attempts of impostors [26] to all the attempts made. It is calculated as `TRR = 1 - FAR`.
- **Receiver- or Relative-Operating Characteristic (ROC):** ROC plot is a visual characterization of trade-off between FAR and TAR [70]. In simple words, it is a plot between true alarms vs. false alarm. The curve is generated by plotting the FAR versus the TAR for varying thresholds to assess classifier's performance [26].

As the parameters are interlinked together, and to avoid redundancy, we report our results in terms of TAR and FAR, and ROC only.

6.3. Results

Table 5 and 6 show the performance of classifiers with full and selected features, respectively. The results are presented for each modality, independently, as well as for binary and ternary feature-level fusion. The performance is measured in terms of TAR averaged for all the 86 users with 120 observations per users using 5-fold cross-validation method.

Table 5: Performance of classifiers with full features for unimodal, bimodal and trimodal configuration based on 5-fold cross-validation.

	Unimodal			Bimodal			Trimodal
Modalities	Swipe	Voice	Face	Voice + Face	Swipe + Voice	Swipe + Face	Swipe + Voice + Face
Total number of features	*33*	*104*	*256*	*380*	*137*	*289*	*393*
Classifier	TAR(%)						
Quadratic SVM	**87.0**	**90.9**	**91.2**	**98.2**	**95.1**	**97.5**	**99.0**
Ensemble Bagged Tree	84.7	88.2	85.0	95.2	94.3	96.6	98.2
Weighted KNN	70.2	85.4	88.7	94.7	90.4	94.1	96.7

Table 6: Performance of classifiers with selected features for unimodal, bimodal and trimodal configuration based on 5-fold cross-validation.

	Unimodal			Bimodal			Trimodal
Modalities	Swipe	Voice	Face	Voice + Face	Swipe + Voice	Swipe + Face	Swipe + Voice + Face
Number of selected features	*11*	*20*	*20*	*40*	*31*	*31*	*51*
Classifier	TAR(%)						
Quadratic SVM	**79.99**	**89.60**	**90.61**	**97.63**	**93.53**	**98.04**	**99.04**
Ensemble Bagged Tree	77.66	86.00	86.72	95.04	91.89	97.08	98.02
Weighted KNN	68.83	86.51	90.71	96.36	90.68	96.93	98.26

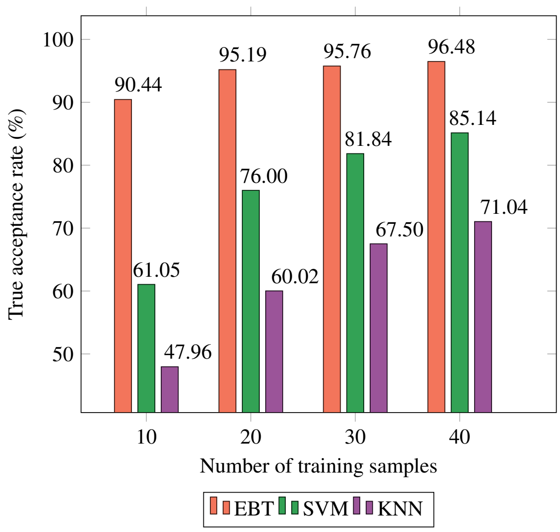


Figure 12: True acceptance rate (TAR) with selected features for trimodal configuration with 10, 20, 30, and 40 training samples.

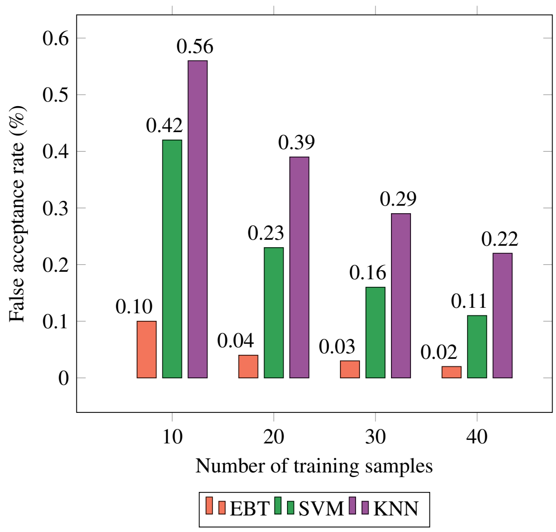


Figure 13: False acceptance rate (FAR) with selected features for trimodal configuration with 10, 20, 30, and 40 training samples.

Figure 12 and 13 show the results of the trimodal system for 10, 20, 30, and 40 training samples with selected feature-set, in term of TAR and FAR, respectively.

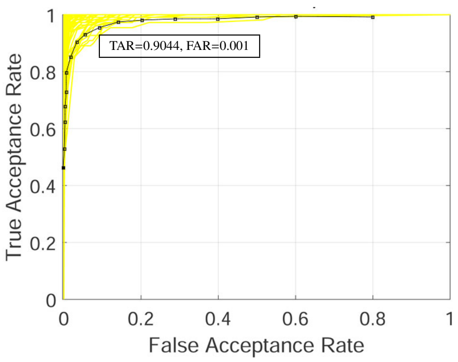
Managing ROC curves for a multi-class classification problem is much more complex in comparison to 2-class classification [70]. Typically, in a multi-class classification model with *n-classes*, the resultant confusion matrix having dimension *n by n* possesses *n* correct classifications (the major diagonal entries) and n^2 =-n possible errors (the off-diagonal entries). According to Fawcett [70], a *class reference formulation* is an efficient method to handle *n-classes* by producing *n*-different *ROC* graphs. Specifically, if C is the set of all classes, ROC graph i reports the classifier performance per class c_i by plotting positive results (P_i), i.e., TAR, as shown in Equation 3 and negative results (N_i), i.e., FAR, as shown in Equation 4.

$$P_i \; = \; c_i \tag{3}$$

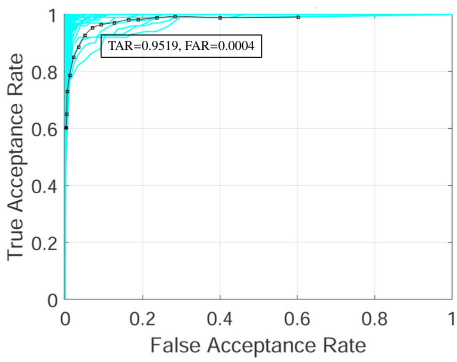
$$N_i \; = \; \bigcup_{j \neq i} c_j \in C \tag{4}$$

This method is reasonably flexible as an optimal threshold t_i can be set, at which TAR is maximum and FAR is minimum. Thus, improving the overall performance of the classification model.

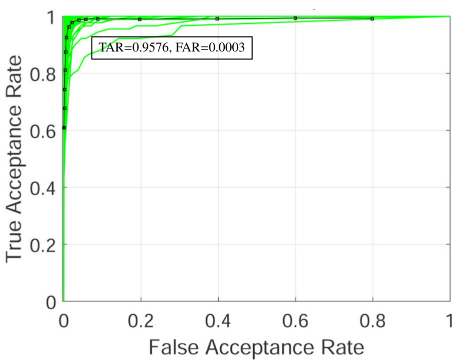
Figure 14 illustrates average ROC curves of EBT classifier for (a) 10, (b) 20, (c) 30, and (d) 40 training samples. In the two-dimensional graphs as shown in Figure 14, TAR is plotted on the Y-axis and FAR is plotted on the X-axis, depicting relative trade-offs between the true positives and false positives. Coordinate $(0, 0)$ represent the strategy of never issuing a positive classification; such a classifier commits no false positive errors but also determines no true positives. However, the opposite strategy, of unconditionally issuing positive classifications, is represented by coordinate $(1, 1)$. Whereas, coordinate $(0, 1)$ represent the perfect classification strategy of



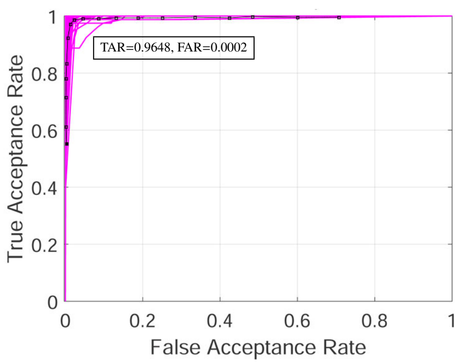
(a) 10 Training Samples



(b) 20 Training Samples



(c) 30 Training Samples



(d) 40 Training Samples

Figure 14: Average ROC curves of EBT classifier for different training samples.

maximizing TAR and minimizing FAR. Readers can observe in Figure 14 with the increase in the number of training samples classifier performance also tends to improve, accordingly.

6.4. Discussion on Results

Cross-validation method is used to evaluate how well the model is trained and how it performs when it is tested on the test dataset. K-fold cross-validation is popular because it is computationally cheap as compared to other cross-validation variants. In K-fold cross-validation, the dataset is divided into K equal folds and the model is trained on the dataset of $K - 1$ folds, and the remaining fold is used to test the system. The process is repeated K times. Cross-validation is preferred when the dataset size is small and it ensures the testing of all the samples. As we had 120 samples for each user, we started the evaluation with 5-fold cross-validation. SVM performed well in this scenario resulting in 99.04% TAR.

Training/Testing split is another method to evaluate the performance of the classifier. The dataset is generally split into two parts, i.e., training and testing sets. The model is trained on the training set (generally, 66% of the whole data) and the remaining test dataset is used to test the model.

Although, Cross-validation method looks justified, because of the low number of observations, however, it seems a bit unrealistic in the real world [71]. In real-world scenarios, e.g., banking applications, generally, the systems require a few attempts to train the classifier and is evaluated everytime the user wants to access their services. Thus, it is worthy to test the classifier with a few numbers of training samples and check for the performance. We tested the pre-trained classifier (trained on 10, 20, 30, and 40 training sample each) and report our obtained results.

In case train/test split scenario, EBT classifier performed better than the SVM and KNN classifiers owing to its ability to reduce the variances and affinity against over-fitting with fewer training samples. It can be noticed that with an increase in the number of training samples, the performance (TAR and FAR) of each classifier improves. For instance, the TAR of EBT classifier improved by +4.75%, +0.57% and +1.29%, whereas FAR became better by −0.06%, −0.01% and −0.01%, with 20, 30 and 40 training samples in comparison to performance with 10 training samples. The same trend can be observed for the other 2 classifiers, i.e., SVM and KNN, in Figure 12 and 13.

7. Conclusions and Future Work

DriverAuth is highly accurate drivers' verification system designed for on-demand ride and ride-sharing services in which customers and the driver-partners are connected to the service provider (server) by the dedicated smartphone applications (clients). Based on the news related to violent altercations, or assaults by malicious drivers and fake drivers offering rides [3, 5, 47, 15]. It is evident that the safety and security of customers are obviously at risk. Therefore, the risk-based verification mechanism can equip service providers to verify the subject at the time of critical decisions (e.g., accepting new registration from a person to join as a driver or assigning new ride assignments to the driver-partners) and trusting the subject with the lives of customers.

We presented a risk-based multi-modal biometric-based driver authentication scheme that uses swipe gesture, voice, and face modalities to profile the driver's identity. We evaluated, DriverAuth, on a dataset of 86 users with 120 observations per user and achieved a TAR of 99.0%, 98.2%, and 96.7% for a trimodal system using SVM, EBT, and KNN classifiers, respectively, on the full feature set.

Feature selection plays a critical role in optimizing the classification model in terms of reduction of feature set dimension and improvement in decision-making time of computationally exhaustive classifiers. We achieved a TAR of 99.04%, 98.02%, and 98.26% using SVM, EBT, and KNN classifiers, respectively, on a selected feature set of dimension 51, which is one-fourth of full feature set, approximately.

In future, we will include the risk-assessment module in DriverAuth to detect and analyze driver-partners' peculiar behaviors or anomalies (e.g., non-professionalism, alcohol-abuse, tiredness, drowsiness, etc.) based on incidents database and driving pattern recordings. We will extend the experimental validation of our proposed scheme on other available datasets, e.g., NIST dataset [72] using advanced machine learning classifiers, e.g., deep learners, in our future work. We will also evaluate and report our scheme's usability and robustness in different attack scenarios.

Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 675320.

References

[1] DMR, "90 amazing uber statistics, demographics and facts." https://expandedramblings.com/index.php/uber-statistics/, 2018. Online web resource.

[2] DMR, "36 interesting lyft statistics and facts." https://expandedramblings.com/index.php/lyft-statistics/, 2018. Online web resource.

[3] J. Horwitz, "Fake drivers and passengers are boosting uber's growth in china." https://qz.com/423288/fake-drivers-and-passengers-are-boosting-ubers-growth-in-china, 2015. Online web resource.

[4] M. Kendall, "Uber slow to boot alleged drunken drivers off the app, state regulators say." https://www.mercurynews.com/2017/04/13/uber-slow-boot-alleged-drunk-drivers-off-app-state-regulators-say/, 2018. Online web resource.

[5] USAtoday, "I got taken for a ride by a fake uber driver. don't become the next victim." https://www.usatoday.com/story/tech/columnist/stevenpetrow/2016/10/12/fake-uber-drivers-dont-become-next-victim/91903508/, 2016. Online web resource.

[6] Whosdrivingyou, "Reported list of incidents involving uber and lyft." http://www.whosdrivingyou.org/rideshare-incidents, 2018. Online web resource.

[7] BBC, "Uber driver background checks not good enough." http://www.bbc.com/news/technology-34002051, 2015. Online web resource.

[8] W. J. A. Al-Nidawi, M. A. Maan, and M. Othman, "Review on national electronic identification system," in *Proceedings of* 4^{th} *International Conference on Advanced Computer Science Applications and Technologies (ACSAT), 2015*, pp. 228–233, IEEE, 2015.

[9] K. Bell, "Uber makes it harder to give drivers bad ratings." https://mashable.com/2017/11/21/uber-makes-it-harder-to-give-bad-ratings/#DghVf5kP0qqR, 2017. Online web resource.

[10] S. Langlois, "Don't tip your uber driver? it could cost you a 5-star rating." https://www.marketwatch.com/story/dont-tip-your-uber-driver-it-could-cost-you-a-5-star-rating-2015-08-12, 2018. Online web resource.

[11] A. Mirsadikov, A. Harrison, and B. Mennecke, "Tales from the wheel: An it-fueled ride as an uber driver," in *Proceedings of Twenty-second Americas Conference on Information Systems*, AMCIS, 2016.

[12] R. Booth, "Uber whistleblower exposes breach in driver-approval process." https://www.theguardian.com/technology/2015/jun/12/uber-whistleblower-exposes-breach-driver-approval-process, 2015. Online web resource.

[13] Uber, "What does the background check include?." https://help.uber.com/h/6970e704-95ac-4ed3-9355-e779a86db366, 2018. Online web resource.

[14] W. Ma, *China's mobile economy: opportunities in the largest and fastest information consumption boom*. John Wiley & Sons, 2016.

[15] J. Bhuiyan, "Uber is facing a class action lawsuit from u.s. riders alleging assault." https://www.recode.net/2017/11/14/16647706/uber-class-action-lawsuit-riders-sexual-assault-rape-violence-background-checks, 2017. Online web resource.

[16] T. Pros and C. O. F. U. Drivers, "Maurice emsellem." https://www.huffingtonpost.com/maurice-emsellem/fingerprinting-uber-drivers_b_10972428.html, 2017. Online web resource.

[17] Uber, "Engineering safety with uber's real-time id check." https://eng.uber.com/real-time-id-check/, 2018. Online web resource.

[18] B. Biggio, G. Fumera, G. L. Marcialis, and F. Roli, "Statistical meta-analysis of presentation attacks for secure multibiometric systems," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 3, pp. 561–575, 2017.

[19] M. I. Gofman, S. Mitra, T.-H. K. Cheng, and N. T. Smith, "Multimodal biometrics for enhanced mobile device security," *Communications of the ACM*, vol. 59, no. 4, pp. 58–65, 2016.

[20] J. Koreman, A. Morris, D. Wu, S. Jassim, H. Sellahewa, J. Ehlers, G. Chollet, G. Aversano, H. Bredin, S. Garcia-Salicetti, *et al.*, "Multi-modal biometric authentication on the securephone pda," 2006.

[21] M. Soltane, N. Doghmane, and N. Guersi, "Face and speech based multi-modal biometric authentication," *International Journal of Advanced Science and Technology*, vol. 21, no. 6, pp. 41–56, 2010.

[22] S. Wang, R. Hu, H. Yu, X. Zheng, R. I. Damper, *et al.*, "Augmenting remote multimodal person verification by embedding voice characteristics into face images," in *Proceedings of International conference on multimedia and expo workshops (ICMEW)*, pp. 1–6, IEEE, 2013.

[23] L. Mezai, F. Hachouf, and M. Bengherabi, "Fusion of face and voice using the dempster-shafer theory for person verification," in *Processings of 7th International Workshop on Systems, Signal Processing and their Applications (WOSSPA)*, pp. 103–106, IEEE, 2011.

[24] T. Kim, H. Park, S. H. Hong, and Y. Chung, "Integrated system of face recognition and sound localization for a smart door phone," *IEEE Transactions on consumer Electronics*, vol. 59, no. 3, pp. 598–603, 2013.

[25] T. Feng, Z. Liu, K.-A. Kwon, W. Shi, B. Carbunar, Y. Jiang, and N. Nguyen, "Continuous mobile authentication using touchscreen gestures," in *Proceedings of IEEE Conference on Technologies for Homeland Security (HST)*, pp. 451–456, IEEE, 2012.

[26] A. Buriro, B. Crispo, F. Del Frari, J. Klardie, and K. Wrona, "Itsme: Multi-modal and unobtrusive behavioural user authentication for smartphones," in *Proceedings of International Conference on Passwords*, pp. 45–61, Springer, 2015.

[27] H. Aronowitz, M. Li, O. Toledo-Ronen, S. Harary, A. Geva, S. Ben-David, A. Rendel, R. Hoory, N. Ratha, S. Pankanti, *et al.*, "Multi-modal biometrics for mobile authentication," in *Proceedings of International Joint Conference on Biometrics (IJCB)*, pp. 1–8, IEEE, 2014.

[28] Z. Akhtar, A. Buriro, B. Crispo, and T. H. Falk, "Multimodal smartphone user authentication using touchstroke, phone-movement and face patterns," in *Proceedings of IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 1368–1372, IEEE, 2017.

[29] A. Buriro, B. Crispo, S. Gupta, and F. Del Frari, "Dialerauth: A motion-assisted touch-based smartphone user authentication scheme," in *Proceedings of the Eighth ACM Conference on Data and Application Security and Privacy*, pp. 267–276, ACM, 2018.

[30] A. Buriro, S. Gupta, and B. Crispo, "Evaluation of motion-based touch-typing biometrics in online financial environments," *BIOSIG 2017*, 2017.

[31] S. C. Eastwood, V. P. Shmerko, S. N. Yanushkevich, M. Drahansky, and D. O. Gorodnichy, "Biometric-enabled authentication machines: A survey of open-set real-world applications," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 2, pp. 231–242, 2016.

[32] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on circuits and systems for video technology*, vol. 14, no. 1, pp. 4–20, 2004.

[33] V. M. Patel, R. Chellappa, D. Chandra, and B. Barbello, "Continuous user authentication on mobile devices: Recent progress and remaining challenges," *IEEE Signal Processing Magazine*, vol. 33, no. 4, pp. 49–61, 2016.

[34] Y. Meng, D. S. Wong, R. Schlegel, *et al.*, "Touch gestures based biometric authentication scheme for touchscreen mobile phones," in *Proceedings of International Conference on Information Security and Cryptology*, pp. 331–350, Springer, 2012.

[35] N. Sae-Bae, N. Memon, K. Isbister, and K. Ahmed, "Multitouch gesture-based authentication," *IEEE transactions on information forensics and security*, vol. 9, no. 4, pp. 568–582, 2014.

[36] H. Xu, Y. Zhou, and M. R. Lyu, "Towards continuous and passive authentication via touch biometrics: An experimental study on smartphones," in *Proceedings of Symposium On Usable Privacy and Security, SOUPS*, vol. 14568–582, pp. 187–198, 2014.

[37] C. McCool, S. Marcel, A. Hadid, M. Pietikäinen, P. Matejka, J. Cernockỳ, N. Poh, J. Kittler, A. Larcher, C. Levy, *et al.*, "Bi-modal person recognition on a mobile phone: using mobile phone data," in *Proceedings of International Conference on Multimedia and Expo Workshops (ICMEW)*, pp. 635–640, IEEE, 2012.

[38] F. Porée, J. Mariéthoz, S. Bengio, and F. Bimbot, "The banca database and experimental protocol for speaker verification," tech. rep., IDIAP, 2002.

[39] S. Garcia-Salicetti, C. Beumier, G. Chollet, B. Dorizzi, J. L. Les Jardins, J. Lunter, Y. Ni, and D. Petrovska-Delacrétaz, "Biomet: A multimodal person authentication database including face, voice, fingerprint, hand and signature modalities," in *Proceedings of International Conference on Audio-and Video-based Biometric Person Authentication*, pp. 845–853, Springer, 2003.

[40] L. Zhang, S. Tan, and J. Yang, "Hearing your voice is not enough: An articulatory gesture based liveness detection for voice authentication," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pp. 57–71, ACM, 2017.

[41] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song, "Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication," *IEEE transactions on information forensics and security*, vol. 8, no. 1, pp. 136–148, 2013.

[42] D. C. Dutt, A. B. Somayaji, and M. J. K. Bingham, "System and method for implicit authentication," Oct. 10 2017. US Patent 9,788,203.

[43] W. Kim, S. Suh, and J.-J. Han, "Face liveness detection from a single image via diffusion speed model," *IEEE transactions on Image processing*, vol. 24, no. 8, pp. 2456–2465, 2015.

[44] K. Patel, H. Han, and A. K. Jain, "Secure face unlock: Spoof detection on smartphones," *IEEE transactions on information forensics and security*, vol. 11, no. 10, pp. 2268–2283, 2016.

[45] Uber, "How to become an uber driver." https://www.uber.com/info/how-to-become-an-uber-driver/, 2018. Online web resource.

[46] Y. Moon, "Uber: changing the way the world moves," *Case, Harvard Business School*, no. 9-316, p. 101, 2015.

[47] Whosdrivingyou, "Fake uber drivers pose real threat." http://www.whosdrivingyou.org/rideshare-incidents, 2017. Online web resource.

[48] I. 9000:2015, "Quality management systems fundamentals and vocabulary." https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v1:en, 2015. Online web resource.

[49] S. Gupta, A. Buriro, and B. Crispo, "Demystifying authentication concepts in smartphones: Ways and types to secure access," *Mobile Information Systems*, vol. 2018, 2018.

[50] S. Eastwood and S. Yanushkevich, "Risk profiler in automated human authentication," in *Proceedings of IEEE Symposium on Computational Intelligence for Engineering Solutions (CIES)*, pp. 140–147, IEEE, 2014.

[51] R. D. Labati, A. Genovese, E. Muñoz, V. Piuri, F. Scotti, and G. Sforza, "Biometric recognition in automated border control: a survey," *ACM Computing Surveys (CSUR)*, vol. 49, no. 2, p. 24, 2016.

[52] K. Lai, S. C. Eastwood, W. A. Shier, S. N. Yanushkevich, and V. P. Shmerko, "Mass evidence accumulation and traveler risk scoring engine in e-border infrastructure," *IEEE Transactions on Intelligent Transportation Systems*, 2017.

[53] S. Gupta, A. Buriro, and B. Crispo, "Driverauth: Behavioral biometric-based driver authentication mechanism for on-demand ride and ridesharing infrastructure," *ICT Express*, 2018.

[54] Uber, "Always the ride you want." https://www.uber.com/en-IT/ride/, 2018. Online web resource.

[55] N. Rogmann and M. Krieg, "Liveness detection in biometrics," in *Proceedings of International Conference of the Biometrics Special Interest Group (BIOSIG)*, pp. 1–14, IEEE, 2015.

[56] L. Zhang, S. Tan, J. Yang, and Y. Chen, "Voicelive: A phoneme localization based liveness detection for voice authentication on smartphones," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1080–1091, ACM, 2016.

[57] S. Marcel, M. S. Nixon, and S. Z. Li, "Handbook of biometric anti-spoofing-trusted biometrics under spoofing attacks, ser," *Advances in Computer Vision and Pattern Recognition. Springer*, 2014.

[58] A. A. Ross, A. K. Jain, and K. Nandakumar, "Information fusion in biometrics," *Handbook of Multibiometrics*, pp. 37–58, 2006.

[59] S. Ma and J. Huang, "Penalized feature selection and classification in bioinformatics," *Briefings in bioinformatics*, vol. 9, no. 5, pp. 392–403, 2008.

[60] D. P. W. Ellis, "Plp, rasta, mfcc, and inversion in matlab." http://www.ee.columbia.edu/~dpwe/resources/matlab/rastamat/, 2005. Online web resource.

[61] J. Kannala and E. Rahtu, "Bsif: Binarized statistical image features," in *Proceedings of 21^{st} International Conference on Pattern Recognition (ICPR)*, pp. 1363–1366, IEEE, 2012.

[62] A. K. Jain, A. A. Ross, and K. Nandakumar, *Introduction to biometrics*. Springer Science & Business Media, 2011.

[63] U. Mahbub, S. Sarkar, V. M. Patel, and R. Chellappa, "Active user authentication for smartphones: A challenge data set and benchmark results," in *Proceedings of 8^{th} International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pp. 1–8, IEEE, 2016.

[64] R. J. Urbanowicz, M. Meeker, W. LaCava, R. S. Olson, and J. H. Moore, "Relief-based feature selection: introduction and review," *Journal of Biomedical Informatics*, vol. 85, pp. 189–203, 2018.

[65] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*.

Morgan Kaufmann, 2016.

[66] L. C. Molina, L. Belanche, and À. Nebot, "Feature selection algorithms: A survey and experimental evaluation," in *Proceedings of International Conference on Data Mining, ICDM*, pp. 306–313, IEEE, 2002.

[67] Matlab, "Classification learner app." https://in.mathworks.com/help/stats/classification-learner-app.html, 2018. Online web resource.

[68] Matlab, "Choose classifier options." https://in.mathworks.com/help/stats/choose-a-classifier.html, 2018. Online web resource.

[69] ISO, "Iso/iec 24713-2:2008(en)." https://www.iso.org/obp/ui/#iso:std:iso-iec:24713:-2:ed-1:v1:en, 2016. Online web resource.

[70] T. Fawcett, "An introduction to roc analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.

[71] A. Buriro, Z. Akhtar, B. Crispo, and S. Gupta, "Mobile biometrics: Towards a comprehensive evaluation methodology," in *Proceedings of International Carnahan Conference on Security Technology (ICCST)*, pp. 1–6, IEEE, 2017.

[72] P. J. Grother, M. L. Ngan, and G. W. Quinn, "Face in video evaluation (five) face recognition of non-cooperative subjects," tech. rep., 2017.

Biography

1.		Sandeep Gupta is a Ph.D. student at the University of Trento, Italy. He received his Master of Technology (M.Tech) in Electronics and communication engineering from Dr. A. P. J. Abdul Kalam Technical University, India. He is the recipient of prestigious Marie Sklodowska-Curie research fellowship. He, previously, worked in the field of information technology with Samsung, Accenture, and Mentor Graphics (now Siemens). His research interests include biometrics, user authentication, risk-based mechanisms on emerging user interfaces, machine learning, and system architecture design.
2.		Attaullah Buriro is currently working as associate professor at KFUEIT Rahim Yar Khan University in Pakistan and holds the postdoc position at DISI Security Lab, University of Trento. He received the Ph.D. degree in Information and Communication Technology (security and privacy) from the University of Trento, Italy, in February 2017. He obtained his Master of Engineering (M.E.) in Telecommunication from NED University of Engineering and Technology, Karachi and Bachelor of Engineering (B.E.) in Electronics from Mehran University of Engineering and Technology, Jamshoro. He has previously worked as an Electronic Engineer in the Pakistan Meteorological Department (July 2007 to May 2017). His research interests include biometrics, authentication, and access control, Internet of Things (IoT), machine learning, artificial intelligence, and data mining
3.		Bruno Crispo received the Ph.D. degree in computer science from University of Cambridge, UK. in 1999, having received the M.Sc. degree in computer science from University of Turin, Italy, in 1993. He is an associate professor at the University of Trento since September 2005. Prior to that, he was associate professor at Vrije Universiteit in Amsterdam. He is the co-editor of the Security Protocol International Workshop proceedings since 1997. He is a member of ACM. His main interests span across the field of security and privacy. In particular, his recent work focuses on the topic of security protocols, access control in very large distributed systems, distributed policy enforcement, embedded devices, and smartphone security and privacy, and privacy-breaching malware detection. He has published more than 100 papers in international journals and conferences on security-related topics.