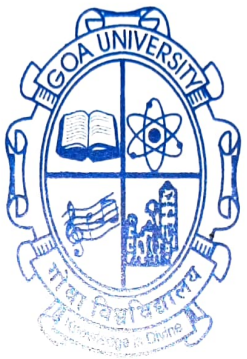
PRESENTATION ATTACK ON SMARTPHONE UNDER SURVEILLANCE SCENARIO



Submitted by

Gauresh N.Naik

Under the Supervisor of

Dr. Narayan Vetrekar

M.Sc. Electronicc

School of Physical and Applied Sciences

A Dissertation Report

Submitted in partial fulfillment of Masters Degree

Course code and CourseTilte :ELE-625 Project

Examined by: Date:

Declaration

I hereby declare that the data presented in this Dissertation / Internship report entitled, "PRESENTATION ATTACK ON SMARTPHONE UNDER SURVEILLANCE SENARIO" is based on the results of investigations carried out by me in the M.Sc. Electronic at the School of Physical and Applied Science, Goa University under the Supervision Dr. Narayan Vetrekar and the same has not been submitted elsewhere for the award of a degree or diploma by me. Further, I understand that Goa University or its authorities will be not be responsible for the correctness of observations / experimental or other findings given the dissertation. I hereby authorize the University authorities to upload this dissertation on the dissertation repository or anywhere else as the UGC regulations demand and make it available to any one as needed.

Mr. Gauresh N. Naik

Seat no: 22P0360007

Electronics Discipline, School of Physical and Applied Science

Date:

Place: Goa University

Certificate

This is to certify that the dissertation report "PRESENTATION ATTACK ON SMARTPHONE UNDER SURVEILLANCE SCENARIO" is a bonafide work carried out by Mr Gauresh N. Naik under my supervision in partial fulfilment of the requirements for the award of the degree of Masters in the Discipline Electronics at the School of Physical and Applied Sciences, Goa University.

Dr. Narayan Vetrekar

Electronics Discipline,

School of Physical and Applied Science

Date: 21/05/2024

Signature of Dean of the School

Date:

Place: Goa University

School Stamp

Acknowledgements

I would like to express my sincerest thanks to everyone who has contributed to this project First and foremost, I would like to thank all the students who have voluntarily participated in this study and have provided me with the milk from their villages giving me their time to help me collect the data. I am grateful to Dr. Narayan Vetrekar my guide for his constant support, assistance, and guidance at every stage, which has been instrumental in the success of this project. I express my deep gratitude towards all the teaching faculty Prof. Rajendra . S. Gad, Prof. Jivan S. Parab, Dr. Marlon Sequeira, Dr. Aniketh Gaonkar and Dr. Sandeep Gawali.

A sincere thank you to Miss Krishna, Miss Marissa, Mrs. Ashwini, Sir. Lopes, Sir Vishant and Sir Ramchandra for their help and support. Their knowledge and suggestions have been crucial in helping me carry out, And execute the project. I would like to express my gratitude to my family and close friends for their unfailing support and inspiration during this Project

Mr. Gauresh N. Naik

Electronics Discipline,

School of Physical and Applied Sciencel

Abstract

The research project, "PRESENTATION ATTACK ONSMARTPHONE UNDER SURVEILLANCE SCENARIO," explores biometric authentication, focusing on face recognition, amidst cybersecurity concerns. It aims to investigate smartphone capabilities in detecting presentation attacks within surveillance contexts. The study involves designing a robust data acquisition protocol, benchmarking detection algorithms, and evaluating performance against 2D and 3D attack scenarios. By leveraging diverse datasets and innovative methodologies, the project aims to advance facial authentication in surveillance applications, addressing critical cybersecurity challenges.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Background

Authentication is the method of verifying the uniqueness of a user or system, confirming that the person or entity trying to access a system or data is who they claim to be. Authentication mechanisms typically involve presenting credentials, such as usernames and passwords, security tokens, biometric data (like fingerprints or facial recognition), or digital certificates [**?**]. It is a fundamental aspect of security in various domains, including computer systems, networks, websites, and applications. Without proper authentication, unauthorized individuals or systems could gain access to sensitive information or resources, leading to security breaches and potential harm. Depending on the context and security requirements, authentication can be achieved using various methods like passwords, Two-Factor Authentication (2FA), Public Key Infrastructure (PKI), OAuth and OpenID Connect, and Token-based Authentication.

The shift towards biometric modalities for authentication represents a response to the limitations and vulnerabilities of traditional authentication methods like passwords and even some advanced methods like Two-Factor Authentication (2FA) or Token-based Authentication. Biometric modalities offer a higher level of security compared to traditional authentication methods, as biometric characteristics such as fingerprints, iris patterns, or facial features are unique to each individual and difficult to replicate or spoof. This makes it harder for unauthorized users to gain

access to systems or data. Additionally, biometric authentication methods are often more convenient for users compared to remembering passwords or carrying physical tokens, improving the overall user experience and reducing the likelihood of insecure practices. Furthermore, biometric authentication reduces the risk of credential theft since biometric data is not easily stolen or intercepted in transit, enhancing accountability and traceability in systems. Moreover, in some industries such as finance and healthcare, regulatory requirements mandate the use of strong authentication methods to protect sensitive data and ensure compliance with privacy regulations, making biometric authentication methods a preferred choice. Overall, the shift towards biometric authentication reflects a desire for stronger security, improved user experience, and compliance with regulatory requirements in an increasingly digital and interconnected world, despite its challenges such as privacy concerns and potential breaches.[3]

Biometric modalities refer to the distinct biological characteristics or traits that are utilized for biometric authentication. These modalities capture unique physiological or behavioral attributes of individuals, which can be used to verify their identity. Common biometric modalities include:

1.1.1 Biometric Modalities

There are different types of biometric modalities based on physiologial and behavioural traits for biometric authentication.

Fingerprint Recognition

Fingerprint recognition is a biometric technique that entails the acquisition and examination of the distinctive ridge and valley patterns on a person's fingertip. It is considered one of the most established and extensively employed biometric modalities owing to the inimitability and constancy of fingerprints. This technology finds its applications in various domains such as law enforcement, access control, mobile device security, and financial transactions.[7]

Iris Recognition

Iris recognition is a biometric technology that involves the analysis of the distinct patterns present in the colored region of the eye, specifically the iris. These patterns are formed during embryonic development and remain stable throughout life, making iris recognition a highly reliable and accurate biometric modality. This technology is particularly well-suited for applications such as surveillance and border control, where its capability to accurately identify individuals from a distance is highly advantageous. Moreover, iris recognition is also applied in access control systems, healthcare for patient identification, and airport security.

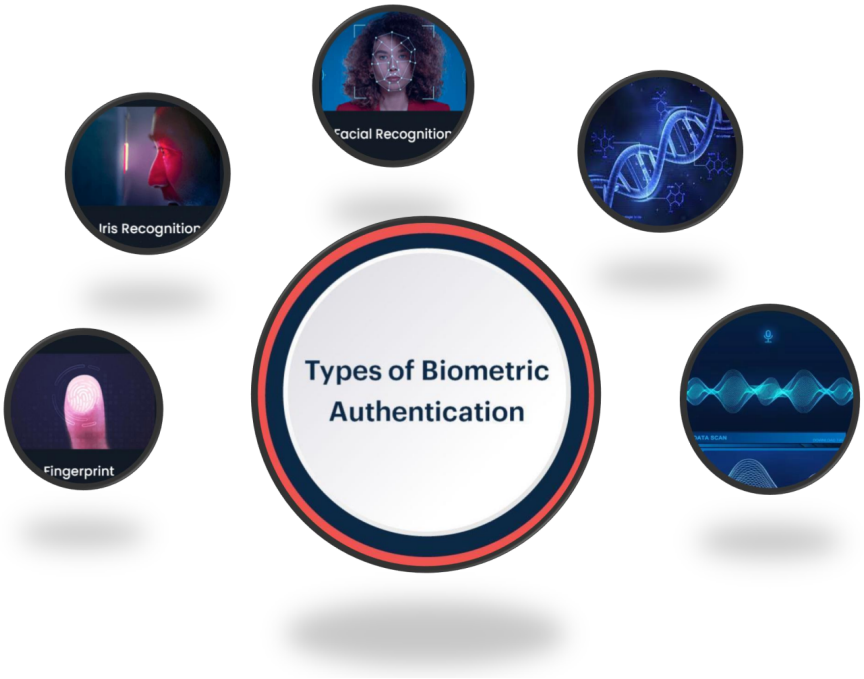


Fig. 1.1 Diverse biometric traits, including fingerprint recognition, iris scanning, facial recognition, voiceprint analysis, hand geometry, retina scanning, and vein pattern recognition, serve as distinct identifiers

Facial Recognition

Facial recognition systems are biometric technologies that operate by analyzing the unique characteristics of an individual's face, such as the interocular distance, nose length, and mouth curvature. This modality finds widespread application in domains such as surveillance, law enforcement, access control, and identity verification on mobile devices. In recent years, facial

recognition algorithms have undergone significant advancements, enabling real-time recognition in varying lighting conditions and angles, which has further enhanced the efficacy and reliability of this technology.

Voice Recognition

Voice recognition is a biometric technology that identifies and analyzes the distinctive features of an individual's voice, including pitch, tone, and cadence. This modality is widely employed for speaker verification in various applications, such as telephone banking, voice-controlled devices, and secure access to systems and services. To further enhance the accuracy of voice recognition systems, behavioral biometrics, such as speech patterns and intonation, may also be incorporated. These advancements have significantly improved the reliability and efficacy of voice recognition technology in recent years.

Hand Geometry

Hand geometry recognition is a biometric technology that captures and analyzes the physical features of an individual's hand, such as the size and shape of the palm and fingers. This modality is commonly employed in access control systems, time and attendance tracking, and industrial environments that require high levels of hygiene. Hand geometry systems are considered relatively simple and cost-effective compared to other biometric modalities, making them a popular choice for various applications.

Retina Recognition

Retina scanning is a biometric technology that involves the acquisition of an image of the blood vessel patterns present at the back of an individual's eye (retina) using infrared light. This modality is highly accurate and resistant to spoofing attacks, as retina patterns are unique to each individual and challenging to replicate. Retina scanning finds its application in high-security environments such as government facilities, military installations, and research laboratories, where its unparalleled accuracy and resistance to fraud are highly advantageous.

Signature Dynamics

Signature recognition systems are biometric technologies that operate by analyzing the distinctive characteristics of an individual's signature, such as stroke sequence and pressure. This modality is widely implemented in banking and financial institutions for the authentication of signatures on checks and documents. Furthermore, it finds application in legal and administrative contexts for identity verification purposes. The uniqueness and consistency of an individual's signature make it an effective biometric modality for a range of applications that require high levels of accuracy and security.

Keystroke Dynamics

Keystroke dynamics is a biometric technology that involves analyzing the unique typing patterns of individuals, such as typing speed, rhythm, and errors. This modality can be employed for continuous authentication on computers and mobile devices, detecting unauthorized access based on typing behavior. It offers a non-intrusive and passive form of authentication, which is highly advantageous in various contexts. However, its accuracy may be influenced by factors such as an individual's typing habits and environmental conditions. Despite these limitations, keystroke dynamics remains an effective barometric modality, offering a unique means of identifying individuals based on their typing patterns.

These biometric modalities vary in terms of accuracy, ease of use, and applicability to different use cases. Organizations often select biometric modalities based on factors such as security requirements, user acceptance, and the specific context of implementation. Additionally, some systems may employ multiple biometric modalities in combination (multimodal biometrics) to enhance accuracy and robustness. But facial biometric-based authentication offers several advantages that make it a compelling choice among biometric modalities:

Facial recognition offers a convenient and seamless authentication experience to users, which eliminates the need for physical tokens or passwords. Facial biometrics are non-intrusive and do not require physical contact with a device or sensor, which makes it a preferred choice for applications where hygiene or user comfort is a concern. The familiarity of facial recognition

to most people due to its widespread use in consumer devices such as smartphones makes it more acceptable to users compared to other authentication methods. Facial recognition systems can provide rapid authentication, which makes it suitable for high-traffic environments such as airports, stadiums, and public transportation hubs. Furthermore, facial biometric authentication is accessible to a wide range of users, including those with disabilities or impairments which may make other authentication methods challenging. Many modern devices such as smartphones, tablets, and laptops are equipped with built-in cameras that can capture facial biometric data, which makes it easy to implement facial recognition without the need for additional hardware. When properly implemented, facial biometric authentication can offer a high level of security as facial features are unique to each individual, and advanced algorithms can detect and prevent spoofing attempts using photos or videos. Finally, facial recognition can be combined with other biometric modalities, such as fingerprint or iris recognition, to create multimodal authentication systems that offer enhanced security and reliability.

In the context of access control, biometric authentication has replaced traditional methods such as keycards or PINs in corporate buildings, government facilities, schools, hospitals, and residential complexes. Similarly, many mobile devices now incorporate biometric authentication features, enhancing device security and allowing users to unlock their devices and access sensitive information using their unique biometric traits. In the banking and financial services industry, biometric authentication is used for identity verification during account opening, transactions, and online banking, while law enforcement agencies use biometric authentication for criminal identification, suspect tracking, and forensic investigations. Biometric authentication is also employed in border control and immigration, healthcare, retail and hospitality, and transportation for various purposes, including enhancing security, improving efficiency, and providing personalized services. These examples demonstrate the diverse range of applications where facial-based authentication is currently being used, highlighting its growing importance in various sectors for enhancing security, efficiency, and user experience[**?**].

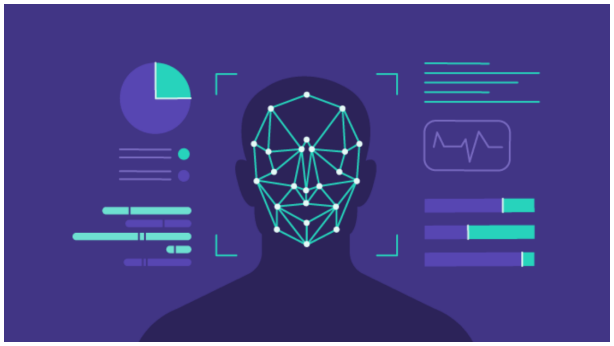


Fig. 1.2 Exploring facial biometrics: Unlocking identity through the unique features of the human face

1.1.2 Facial Biometric Challenges

However, despite the widespread adoption of facial-based authentication, challenges persist, particularly regarding presentation attacks. Presentation attacks involve malicious attempts to deceive the facial recognition system by using counterfeit or manipulated biometric data, such as photographs, videos, or masks, to gain unauthorized access. These attacks pose a significant threat to the integrity and reliability of facial biometric authentication systems, potentially compromising security across various sectors.

These attacks, also referred to as spoofing attacks or biometric attacks, pose significant security risks and are categorized into different types based on the nature of the Presentation Attack Instrument (PAI) employed.

Photo attacks represent one of the most prevalent and concerning types of presentation attacks. They involve presenting a photograph of the targeted individual's face to the facial recognition system's sensor. Photo attacks are particularly worrisome due to their simplicity and accessibility. Attackers can easily obtain high-quality face images from social media platforms or covertly capture them using digital cameras, making them a potent threat to facial recognition systems unless robust countermeasures are implemented.

Furthermore, 2D presentation attacks exploit the use of two-dimensional representations of genuine faces, including printed photographs or digital images displayed on screens. While relatively straightforward to execute, these attacks can bypass authentication mechanisms if facial

recognition systems lack effective anti-spoofing measures capable of distinguishing between genuine and fake images.

In contrast, 3D presentation attacks involve presenting three-dimensional representations of genuine faces to facial recognition systems. These attacks may include sophisticated methods such as 3D-printed masks, sculpted facial prosthetics, or computer-generated 3D models. Their complexity and ability to replicate the depth and structure of real faces make them challenging to detect, especially for facial recognition systems without advanced anti-spoofing capabilities.[1]

To mitigating the risks associated with presentation attacks necessitates the implementation of robust anti-spoofing measures, including liveness detection techniques, multi-modal biometrics, and continuous authentication mechanisms. By proactively addressing vulnerabilities and staying abreast of emerging threats, organizations can enhance the security and reliability of their facial biometric authentication systems in the face of evolving presentation attack techniques. As such, ongoing research and development efforts are focused on advancing anti-spoofing techniques and robustness measures to detect and mitigate presentation attacks effectively. Addressing these concerns is crucial to maintaining trust in facial-based authentication systems and ensuring their continued effectiveness in enhancing security, efficiency, and user experience across diverse applications.

In response to the imperative of mitigating risks associated with presentation attacks, our research endeavors have focused on implementing robust anti-spoofing measures, leveraging the capabilities of smartphone sensors to record both 2D and 3D facial data. Through systematic experimentation with various types of attacks, we have meticulously collected data using different smartphone sensors to develop robust algorithms for presentation attack classification. As in today's world, smartphones have become a ubiquitous tool for data collection in various applications. This widespread use of smartphones can be attributed to the convergence of several enabling technologies, such as high-resolution cameras, powerful processors, and advanced sensors, which make it easier to collect data on the go. Additionally, smartphones offer inherent advantages such as portability, accessibility, and user-friendliness, making them an ideal platform

for data collection in different fields. As a result, researchers and developers are increasingly leveraging smartphones to collect data for various applications, including health monitoring, environmental monitoring, and social research.

1.2 Image capturing tech is for taking digital images

Fig. 1.3 image acquisition technology

1.2.1 Digital Cameras

Digital cameras are perhaps the most common devices used for capturing digital images. They come in various forms, including compact cameras, DSLRs (Digital Single-Lens Reflex), mirrorless cameras, and action cameras. These cameras use image sensors (such as CCD or CMOS) to convert optical images into digital signals.

1.2.2 Smartphone Cameras

Smartphone cameras have become ubiquitous, offering increasingly sophisticated imaging capabilities. They typically use CMOS image sensors along with lenses and software processing to capture and enhance images. Features such as multiple lenses (wide-angle, telephoto, ultra-wide), computational photography, and AI-based enhancements have become common in smartphone cameras.

1.2.3 Webcams

Webcams are cameras built into computers or external devices primarily used for video calling, conferencing, or live streaming. They typically use CMOS sensors and connect to computers via USB or other interfaces.

1.2.4 Action camera

Action cameras are rugged, compact cameras designed for capturing dynamic activities such as sports, adventures, and extreme sports. They often feature wide-angle lenses, waterproof housings, and shockproof designs.

1.2.5 Drones

Drones (unmanned aerial vehicles) are equipped with cameras for aerial photography and videography. Drone cameras range from small, integrated units on consumer drones to high-end professional cameras mounted on professional-grade drones.

1.2.6 Security Cameras

Security cameras, also known as surveillance cameras, are used for monitoring and recording activities in various environments. They may include features such as motion detection, night vision, and remote monitoring capabilities.

1.2.7 Machine Vision Cameras

Machine vision cameras are specialized cameras used in industrial applications for tasks such as quality control, inspection, and robotic guidance. They often feature high-speed, high-resolution sensors optimized for specific applications.

1.2.8 Thermal Cameras

Thermal cameras capture images based on heat signatures rather than visible light. They are used for applications such as surveillance, search and rescue, industrial inspection, and medical diagnostics.

1.2.9 360-Degree Cameras

360-degree cameras capture immersive, panoramic images and videos, allowing viewers to explore the entire scene. They often use multiple lenses and sensors to capture a full spherical view.

1.3 Smartphone based Face Biometric

Smartphones are equipped with an array of integrated sensors, including high-resolution cameras, microphones, GPS receivers, accelerometers, and gyroscopes, which facilitate the capture of diverse data types, including biometric data such as facial images, fingerprints, and voice samples. The portability and mobility of smartphones make them ideal for data collection in various environments and situations. Their compact form factor allows users to carry them effortlessly, enabling data collection on the go, whether in urban settings, remote locations, or while traveling.

Smartphones feature intuitive user interfaces, touchscreens, and mobile applications that streamline the data collection process. These interfaces can guide users through data capture steps, prompt for inputs, and provide real-time feedback, enhancing user engagement and data quality. Additionally, smartphones are inherently connected devices, capable of wireless communication via cellular networks, Wi-Fi, and Bluetooth. This connectivity enables real-time data transmission, synchronization with cloud services, and remote access to data collection platforms, facilitating seamless collaboration and data sharing.

Modern smartphones boast powerful processors, ample storage, and advanced graphics capabilities, enabling them to handle complex data processing tasks locally. They can perform

on-device analysis, image processing, and machine learning algorithms, reducing reliance on external computing resources and enhancing data privacy. Moreover, smartphones offer robust security features to safeguard user data, including biometric authentication methods such as fingerprint scanners and facial recognition. Biometric data collected on smartphones can be encrypted, stored securely, and protected from unauthorized access, ensuring user privacy and data integrity.

Leveraging smartphones for data collection is cost-effective compared to investing in dedicated data collection devices or specialized equipment. With widespread adoption and competitive pricing, smartphones provide a cost-efficient solution for organizations seeking to scale data collection efforts while minimizing hardware expenses. Furthermore, continuous advancements in smartphone technology, including improvements in sensor capabilities, processing power, battery life, and connectivity options, further enhance their suitability for data collection applications. These advancements enable innovative use cases and expand the possibilities for data-driven insights across industries

Despite the many advantages of utilizing smartphones for data collection, there are also several challenges that must be considered. One of the most significant challenges is the limited battery life of smartphones, which can be drained quickly by continuous data collection. This necessitates frequent recharging or the use of external battery packs to ensure uninterrupted data collection. Another challenge is the finite internal storage capacity of smartphones, which can pose difficulties when collecting large data volumes, such as high-resolution images or videos. Efficient storage management or the offloading of data to external storage or cloud services become necessary to overcome this challenge. Smartphones are also susceptible to security threats such as malware and unauthorized access, which can compromise collected data. Implementing robust security measures, such as encryption and secure data transmission protocols, is essential to mitigate these risks. Privacy concerns can also arise when collecting sensitive data on smartphones, necessitating adherence to privacy regulations, obtaining user consent, and transparently communicating data collection practices to build user trust. The quality and accuracy of collected data can be compromised by environmental conditions, sensor

limitations, or technical glitches. Implementing quality control measures, such as data validation and error detection, can help ensure data reliability. Reliance on network connectivity for data transmission can also be disrupted by poor coverage or network outages, requiring offline data collection capabilities or alternative communication methods to address connectivity challenges. Additionally, the wide variety of smartphone models and software versions can lead to compatibility issues and software fragmentation, requiring careful consideration during application development. Encouraging user engagement and compliance with data collection protocols can also be challenging due to privacy concerns or lack of incentive. Designing user-friendly interfaces and offering incentives can improve participation rates and enhance data collection quality.

Chapter 2

Literature Review and Contribution

2.1 Literature Review

It seems that the study titled "Deeply vulnerable: a study of the robustness of face recognition to presentation attacks" by A Mohammadi, S Bhattacharjee, and S Marcel, published in Iet Biometrics in 2018, discusses the vulnerability of deep-learning-based face-recognition (FR) methods to presentation attacks (PA). It highlights that while FR methods based on deep neural networks (DNN) have shown significant improvements in recognition performance, a trustworthy face-verification system should also be capable of resisting various kinds of attacks, including PA. The study shows that DNN-based FR systems tend to score bona fide and PA samples similarly, making them extremely vulnerable to PAs. The experiments conducted in the study indicate that the vulnerability of the studied DNN-based FR systems is consistently higher than 90%, and often higher than 98%.[17]

The paper addresses the lack of reliable software-based face presentation attack detection (PAD) methods for mobile authentication. It notes that existing datasets cover various attack scenarios but lack standardized evaluation protocols for assessing generalization capabilities across different conditions.To fill this gap, the authors introduce the OULU-NPU database. It aims to evaluate PAD methods across realistic mobile authentication scenarios, considering unknown environmental conditions, acquisition devices, and presentation attack instruments.

The database includes 5940 videos of 55 subjects in three environments, recorded using six different smartphones. High-quality print and video-replay attacks were created using various printers and display devices.The database features four evaluation protocols, each introducing previously unseen conditions to the test set. This enables fair comparisons of generalization capabilities. Baseline results using color texture analysis-based PAD method highlight the database's challenges.[8] . The REPLAY-MOBILE database, introduced in the 2016 International Conference of the Biometrics Special Interest, aims to provide robust countermeasures for face presentation-attack detection (PAD) on mobile devices. The database includes 1,200 videos of 40 clients, containing both genuine videos and various types of presentation attacks. It also provides three separate sets for training, validating, and testing classifiers for the face-PAD problem, making it easier for researchers to compare new approaches with existing algorithms in a standardized way. The database also offers baseline results using state-of-the-art approaches based on image quality analysis and face texture analysis.[10]

The paper "Spoofing attacks to 2D face recognition systems with 3D masks" by N. Erdogmus and S. Marcel, published in 2013, highlights the vulnerability of 2D face recognition systems to spoofing attacks using 3D facial masks. The study aims to examine possible 3D attack instruments and assess the spoofing performance for each type of mask. A small database with six different types of 3D facial masks is constructed and used to conduct experiments on state-of-the-art 2D face recognition systems. The study shows that 2D face recognition systems are vulnerable to spoofing attacks using 3D facial masks and highlights the need for developing robust countermeasures to detect and prevent such attacks.[11]

The paper "Presentation attack detection for face recognition using light field camera" by R. Raghavendra, K.B. Raja, and C. Busch, published in IEEE Transactions on Image Processing in 2015, proposes a novel approach for face presentation attack detection using a light field camera (LFC). The LFC records the direction of each incoming ray, which renders multiple depth or focus images in a single capture. The proposed approach involves exploring the variation of focus between multiple depth images rendered by the LFC to reveal presentation attacks. A new face artefact database is collected using LFC, comprising of 80 subjects, and extensive

experiments carried out on the light field face artefact database have revealed the outstanding performance of the proposed PAD scheme when benchmarked with various well-established state-of-the-art schemes.[19]

The paper "Face liveness detection with component dependent descriptor" by J. Yang et al., presented at the 2013 International Conference on Biometrics, proposes a component-based face coding approach for liveness detection to detect spoofing attacks. The method involves locating face components, coding low-level features for each component, deriving high-level face representation, and concatenating histograms for identification. The proposed framework utilizes micro differences between genuine and fake faces while retaining inherent appearance differences among different components and achieves the best liveness detection performance in three databases.[23]

The paper presents a novel software-based fake detection method that uses image quality assessment to enhance the security of biometric recognition systems. The proposed approach is fast, user-friendly, and non-intrusive, and can be used to detect different types of fraudulent access attempts in multiple biometric systems. The method uses 25 general image quality features extracted from the same image acquired for authentication purposes to distinguish between legitimate and impostor samples. The experimental results show that the proposed method is highly competitive compared with other state-of-the-art approaches and that the analysis of the general image quality of real biometric samples reveals highly valuable information that can be efficiently used to discriminate them from fake traits. The study focuses on fingerprint, iris, and 2D face recognition systems.[13]

The paper "Face anti-spoofing with multifeature videolet aggregation" by T.A. Siddiqui et al., presented at the 2016 International Conference on Pattern Recognition, proposes a novel multi-feature evidence aggregation method for face spoofing detection. The method fuses evidence from features encoding texture and motion properties in the face and surrounding scene regions, using local binary pattern and motion estimation algorithms. The multi-feature windowed videolet aggregation of these orthogonal features, coupled with support vector machine-based

classification, provides robustness to different attacks. The proposed approach is evaluated on three standard public databases and achieves an equal error rate of 3.14%, 0%, and 0% on .[22]

The paper "Generalized face anti-spoofing by detecting pulse from face videos" by X. Li et al., presented at the 2016 International Conference on Pattern Recognition, proposes a robust anti-spoofing method by detecting pulse from face videos to differentiate genuine faces from fake ones. The method is based on the fact that a pulse signal exists in a real living face but not in any mask or print material, making it a generalized solution for face liveness detection. The proposed method is evaluated on a 3D mask spoofing database and cross-database experiments with high-quality masks show that the pulse-based method is able to detect even the previously unseen mask type, whereas texture-based methods fail to generalize beyond the development data. Finally, the authors propose a robust cascade system combining two complementary attack-specific spoof detectors, utilizing pulse detection against print attacks and color texture analysis against video attacks.[16]

The paper "Biometric antispoofing methods: A survey in face recognition" by J. Galbally et al., published in IEEE Access in 2014, provides a comprehensive overview of the work carried out over the last decade in the field of antispoofing, with special attention to the face modality. Spoofing, or presentation attack, is a biometric vulnerability where a synthetic or forged version of a genuine biometric trait is presented to the sensor to fool the system into recognizing an illegitimate user as a genuine one. The paper covers theories, methodologies, state-of-the-art techniques, evaluation databases, and provides an outlook into the future of this active field of research.[12]

This paper evaluates various face presentation detection (PAD) techniques to distinguish real face samples from spoof artifacts in mobile scenarios, including photo print and video replay attacks. The study compares the detection performance of 30 representative face PAD methods on three public mobile spoofing datasets and tests the generalization ability of existing methods under cross-database testing scenarios. The paper provides insights to promote both academic research and practical applications.

This paper develops a multi-modal biometric authentication system for smartphones using fingerprint, face, and voice recognition. It proposes a standardized evaluation methodology to assess the system's performance, including a detailed description of the biometric database created for evaluation. The dataset includes diverse populations from different geographic locations, and specific protocols were followed during data acquisition to maintain consistency and quality. The paper also reports the performance evaluation of baseline biometric verification and presentation attack detection on the dataset, allowing for the development of novel algorithms for biometric authentication.[15]

It seems like the paper is about detecting presentation attacks in face biometric systems on smartphones using raw sensor data. The authors proposed a novel approach involving subtracting noise from raw data and computing energy values for detection. They evaluated the proposed method using a newly collected database of 390 live presentation attempts of face characteristics and 1530 attack presentations on the iPhone 6S smartphone. The results showed a lower average classification error achieved, and the proposed method using raw sensor data effectively detects presentation attacks. Overall, the average classification error is significantly lower with the presented approach.[21]

It's interesting to see the focus on Face Anti-spoofing (FAS) in long-distance scenarios like station squares and parks. The introduction of the Surveillance High-Fidelity Mask (SuHiFiMask) dataset seems like a significant step in evaluating algorithms' robustness under quality changes in surveillance scenarios. The face presentation attack detection challenge using this dataset also sounds like a great initiative to assess algorithms for detecting attacks in long-range surveillance scenarios. It's impressive to see the support from various grants and funding sources for this work, including the National Key Research and Development Plan, Chinese Academy Sciences, Chinese National Natural Science Foundation Projects, Science and Technology Development Fund of Macau Project, and the InnoHK program.[1]

Alireza Sepas-Moghaddam's research is focused on using light field imaging technology to improve biometric recognition and presentation attack detection systems. The GUC Light Field Face Artefact Database (GUC-LiFFAD) and IST Lenslet Light Field Face Spoofing Database (IST

LLFFSD) are two databases used for benchmarking and testing proposed solutions. The proposed solution based on a histogram of oriented gradients descriptor showed superior performance when compared to other state-of-the-art alternatives. Furthermore, the importance of using color information was highlighted, as it led to better performance compared to using grayscale information alone. Alireza Sepas-Moghaddam's research focuses on utilizing advancements in light field imaging technology to develop biometric recognition and presentation attack detection systems with improved performance. To facilitate this research, two databases, GUC Light Field Face Artefact Database (GUC-LiFFAD) and IST Lenslet Light Field Face Spoofing Database (IST LLFFSD), have been created that include various artefact types, greyscale and RGB rendered face images, depth maps, and raw light field imaging data in Lytro Light Field Raw (LFR) format.The proposed solution for face presentation attack detection, which is based on a histogram of oriented gradients descriptor, demonstrated outstanding effectiveness and stability, surpassing state-of-the-art alternatives. Additionally, the importance of incorporating color information in benchmarking solutions was emphasized, as the use of color led to performance improvements when compared to using only grayscale information.[4]

The challenges in implementing anti-spoofing techniques for face recognition systems in real scenarios, focusing on generalization, usability, and performance issues, are crucial aspects highlighted in the paper. The lack of publicly available collaborative face-PAD datasets indeed hinders the comparison of anti-spoofing methods in an open and reproducible framework. It's great to see the growing interest in Presentation Attack Detection (PAD) methods, leading to the development of various research works, anti-spoofing databases, and competitions for evaluating new PAD algorithms. The emphasis on analyzing cross-domain performance, database limitations, and usability aspects in face antispoofing is essential for the advancement of this field. Additionally, the need for larger and more representative datasets to improve the performance of current face-PAD methods in real-world scenarios is well noted.[5]

The research paper "Fusion Methods for Face Presentation Attack Detection" by Faseela Abdullakutty, Pamela Johnston, and Eyad Elyan aims to improve the security of face recognition systems by combining deep learning features with traditional color and texture features to detect

face presentation attacks. The study used the Replay Attack dataset to train pre-trained and custom CNN models for binary classification. They employed a feature-fusion method that combines pre-trained deep learning models with traditional color and texture features to enhance the detection of face presentation attacks.The research showed that the fusion methods improved detection rates on public datasets like CASIA, Replay Attack, and SiW by enriching the feature space. The results were reported in various performance metrics, including accuracy, HTER, precision, recall, F1 score, FPR, and FNR, demonstrating the effectiveness of the fusion strategies for detecting face presentation attacks.[2]

It seems like you are sharing information about a research paper titled "Secure Face Unlock: Spoof Detection on Smartphones" written by Keyurkumar Patel, Hu Han, and Anil K. Jain. The paper discusses the development of a face spoof detection system on Android smartphones to analyze image distortion for print and replay attacks. The researchers used the MSU USSA database containing over 1000 subjects for spoof detection, and print and replay attacks were captured using Nexus 5 smartphone cameras. The paper also discusses image distortion analysis, including surface reflection, moire pattern, and color distortion. The proposed approach was found to be effective in detecting face spoofing attacks for print and replay attacks in real application scenarios. The researchers also developed an unconstrained smartphone spoof attack database (MSU USSA) with 1000+ subjects. The system was found to be successful in detecting face spoofing attacks in real application scenarios during user studies.[18]

It's interesting to see the focus on Face Anti-spoofing (FAS) in long-distance scenarios like station squares and parks. The introduction of the Surveillance High-Fidelity Mask (SuHiFiMask) dataset seems like a significant step in evaluating algorithms' robustness under quality changes in surveillance scenarios. The face presentation attack detection challenge using this dataset also sounds like a great initiative to assess algorithms for detecting attacks in long-range surveillance scenarios. It's impressive to see the support from various grants and funding sources for this work, including the National Key Research and Development Plan, Chinese Academy Sciences, Chinese National Natural Science Foundation Projects, Science and Technology Development Fund of Macau Project, and the InnoHK program.[20]

The article "A Smart Spoofing Face Detector by Display Features Analysis" by ChinLun Lai and ChiuYuan Tai likely presents a novel approach to spoofing detection in facial recognition systems. By analyzing display features, the authors propose a method for detecting spoofing attempts, enhancing the security of face recognition systems.The term "smart spoofing face detector" suggests an intelligent system capable of distinguishing between genuine facial images and spoofed ones.Facial biometric authentication systems are susceptible to spoofing attacks using photos, videos, or masks. Research has extensively explored feature extraction using color spaces, such as YCbCr and HSV, to improve accuracy in spoofing attack detection. Texture analysis in facial recognition tasks has also utilized various color spaces. The extracted color features are then input into classification models like ResNet50, VGG16, and MobileNetV2 to effectively detect spoofing attacks. In experimental results, the proposed approach achieved a promising result with the lowest Equal Error Rate (EER) of 3.62% on the CASIA dataset, demonstrating the method's effectiveness.[1]

2.2 Research Qustions

These questions have been formulated based on the findings of previously conducted literature reviews in the field of biometric authentication and identification systems. The questions aim to explore various aspects related to the effectiveness and vulnerabilities of facial recognition technology, including deep-learning-based systems, software-based face presentation attack detection methods, 2D face recognition systems, light field cameras, genuine face characteristics for liveness detection, multi-feature evidence aggregation methods, pulse detection, trends and advancements in biometric antispoofing methods, face presentation attack detection techniques in mobile scenarios, and the integration of multiple biometric modalities for improved security and reliability of authentication systems on smartphones.

- How do deep-learning-based face recognition systems perform when subjected to presentation attacks, and what are the vulnerabilities associated with these systems?

- What are the limitations of existing software-based face presentation attack detection methods for mobile authentication, and how can these limitations be addressed?
- How do 2D face recognition systems respond to spoofing attacks using 3D facial masks, and what are the implications for biometric security?
- What are the advantages of using light field cameras for face presentation attack detection, and how do these cameras compare to traditional methods in terms of accuracy and reliability?
- What are the key features and characteristics of genuine faces that distinguish them from fake ones in biometric authentication systems, and how can these features be effectively leveraged for liveness detection?
- How do multi-feature evidence aggregation methods enhance the robustness of face spoofing detection algorithms, and what are the implications for real-world applications?
- What are the advantages and limitations of using pulse detection as a generalized solution for face liveness detection, and how does it compare to traditional texture-based methods?
- What are the current trends and advancements in biometric antispoofing methods for face recognition, and what are the challenges that need to be addressed for further improvement?
- How do various face presentation attack detection techniques perform in mobile scenarios, and what factors contribute to their effectiveness and generalization capabilities?
- How does the integration of multiple biometric modalities (e.g., fingerprint, face, voice) improve the security and reliability of authentication systems on smartphones, and what are the key considerations for system evaluation and benchmarking?

2.3 Objectives

Facial data acquisition plays a crucial role in surveillance setups, and a well-defined protocol is essential to ensure accurate and reliable data collection. For indoor conditions, the installation

of high-quality surveillance cameras at strategic positions and proper lighting arrangements are crucial to capture facial features accurately. Similarly, for outdoor scenarios, weather-resistant cameras and adequate lighting are necessary to ensure effective facial recognition.

To enhance the precision of facial recognition in surveillance, various experiments are being conducted to evaluate the performance of smartphone models in detecting face mask attacks under diverse scenarios. The experiments aim to test the accuracy, speed, and reliability of the smartphones in detecting face mask attacks in indoor, outdoor, well-lit, and low-light environments. The findings from the experiments will help identify the most effective smartphone models and scenarios for detecting face mask attacks.

Furthermore, a benchmarking study is underway to assess the performance of presentation attack detection algorithms for accurate facial recognition in surveillance. The study involves evaluating various 2D and 3D presentation attack detection algorithms and assessing their accuracy, speed, and reliability in detecting presentation attacks. The results of the study will help identify the most effective presentation attack detection algorithms for accurate facial recognition in surveillance.

These studies are crucial in developing more effective and reliable surveillance systems while ensuring ethical data collection practices. The results from these experiments and benchmarking studies will provide valuable insights into enhancing the precision and reliability of facial recognition systems.

The summary of objective of the project is to improve the performance of facial recognition technology by addressing its limitations and vulnerabilities in various scenarios

- Design a protocol for acquiring facial data in surveillance setups, addressing indoor/outdoor conditions and privacy.
- Conduct experiments to assess smartphone performance in detecting face mask attacks in varied scenarios.
- Benchmark presentation attack detection algorithms for accurate facial recognition in surveillance, including 2D and 3D attacks.

2.4 Contribution

In this project, we aimed to evaluate the performance of different smartphones against face presentation attacks, and construct a comprehensive database of different types of presentation attacks in controlled and uncontrolled environmental conditions. To accomplish this, we collected a large dataset of face images captured in various conditions, including indoor and outdoor settings, and under different lighting conditions. We then simulated different types of presentation attacks, including print attacks, silicon mask attacks,latex mask attack, joker mask attacks, and so on.And collected data on the devices' responses to these attacks. We used this data to construct a database of presentation attacks and bonafide face images, which we then used to train and test different classifier models based on deep learning and conventional classification methods. Our extensive qualitative and quantitative experimental results showed that deep learning models were more effective than conventional classification methods in detecting face presentation attacks, with higher accuracy and lower false positive rates. We also found that the performance of different smartphones varied widely, with some devices showing high accuracy in detecting presentation attacks, while others were more vulnerable to attacks. Overall, our findings suggest that it is important for smartphone manufacturers to continue improving their face recognition technology and security features to better protect users' personal information and data against face presentation attacks

The summary of the contribution of the project.

- Design a protocol for acquiring facial data in surveillance setups, addressing indoor/outdoor conditions and privacy.
- Construct a database of presentation attacks and bonafide face images in various lightening conditions, for of presentation attacks.
- Evaluate the performance of different smartphones against face presentation attacks.
- Training and testing of different classification models.

Chapter 3

Database

3.1 Smartphone Face Database for Presentation Attack

When it comes to collecting data for face spoofing detection, researchers often prioritize controlled indoor environments due to the greater control they have over environmental factors such as lighting, background clutter, and camera angles. This type of controlled environment facilitates the gathering of consistent and high-quality data for training and evaluating the spoofing detection system. However, outdoor data collection poses additional challenges due to the unpredictability of environmental factors such as varying lighting conditions, diverse backgrounds, and potential occlusions. These factors have the potential to impact the performance of the spoofing detection system. Therefore, it is crucial for researchers to collect data from both indoor and outdoor settings to assess the robustness of their spoofing detection algorithm across different environmental conditions. This comprehensive approach is essential to ensure that the system can effectively detect spoofing attempts in real-world scenarios, where environmental factors may vary significantly. Furthermore, this approach enables researchers to identify and address any limitations or biases in the training data, ultimately enhancing the reliability and generalizability of the detection system.

In addition to collecting data from indoor and outdoor settings, researchers can also benefit from gathering data across a diverse range of subjects, spoofing attack types, and devices. By doing so, researchers can ensure that their spoofing detection system is capable of detecting a wide range of potential attacks, including those performed with different devices and by individuals with varying physical characteristics. Moreover, it is important to consider the ethical implications of data collection for face spoofing detection. As biometric data is sensitive personal information, it is crucial to obtain informed consent from participants and ensure that their privacy is protected. Additionally, researchers should take measures to prevent their data from being misused or exploited by third parties. In summary, we can say collecting data for face spoofing detection requires a comprehensive approach that considers not only the environmental conditions but also the range of potential attacks and the ethical implications of data collection. By doing so, researchers can develop a more robust and reliable spoofing detection system that is capable of detecting a wide range of potential attacks in real-world scenarios while ensuring the privacy and security of individuals' biometric data. Our research methodology involved capturing data both indoors and outdoors to comprehensively assess the effectiveness of our presentation attack detection system. By recording data in varied environments, we aimed to develop a robust system capable of accurately detecting presentation attacks across different scenarios, which can ensure thorough testing and validation of our system's performance under real-world conditions.

3.1.1 Indoor Acquisition

When collecting data for smartphone-based spoofing detection, indoor conditions are typically preferred as they provide a controlled environment that mimics everyday scenarios encountered by users. The following key considerations should be taken into account when selecting indoor environments: Firstly, consistent and adequate lighting conditions should be maintained to avoid extreme contrasts or shadows that could impact the quality of facial images captured by the smartphone camera. Secondly, backgrounds that are typical of indoor settings should be chosen, such as walls, furniture, or indoor decorations, while avoiding overly cluttered or

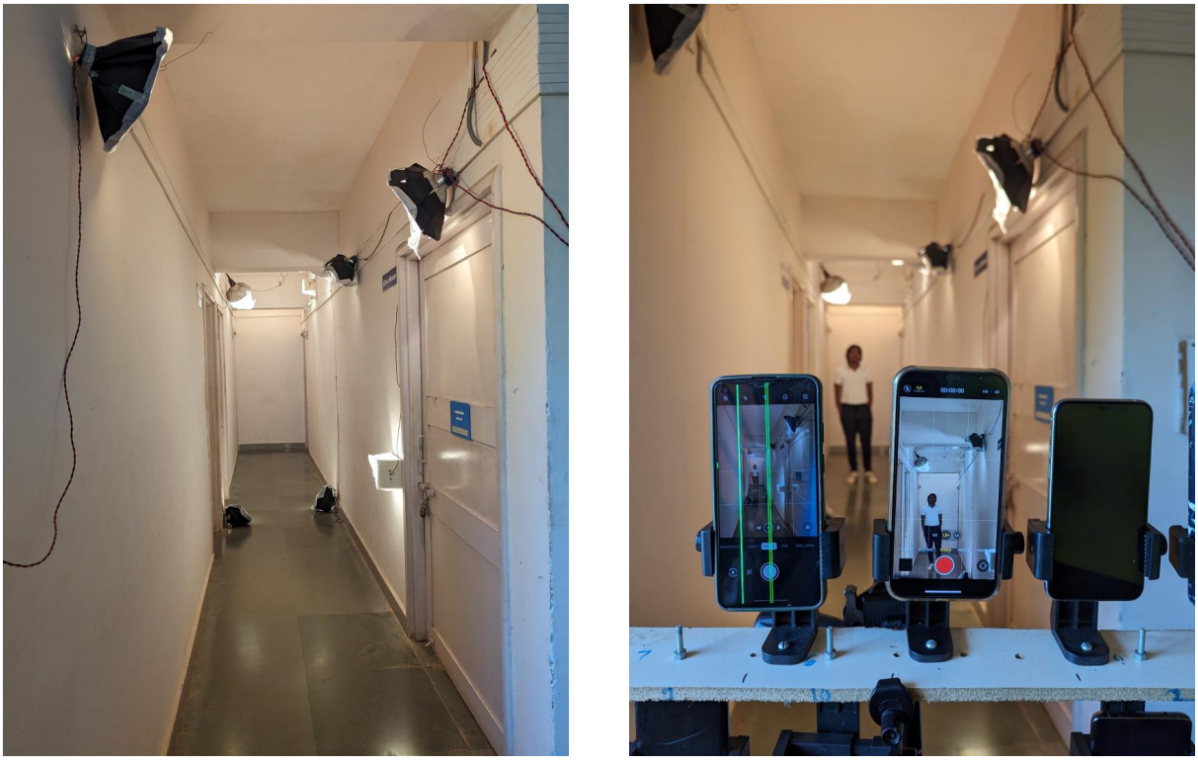


Fig. 3.1 Indoor environment setup for experimentation or observation

visually complex backgrounds that could distract from the facial features being captured. Thirdly, different camera angles and distances should be experimented with to capture a variety of facial poses and expressions, allowing the spoofing detection algorithm to recognize genuine facial images from different perspectives. Fourthly, participants should be encouraged to display a range of facial expressions during data collection, including neutral, smiling, and frowning expressions, to improve the robustness of the spoofing detection system against attempts to mimic specific facial gestures. Furthermore, it is important to maintain consistent environmental conditions throughout the data collection process to ensure reliable and reproducible results, while minimizing external factors such as noise, temperature fluctuations, or interruptions that could affect data quality. Including participants from diverse demographic backgrounds (e.g., age, gender, ethnicity) is important to ensure the spoofing detection system is inclusive and effective for a wide range of users. Finally, ethical considerations are crucial when collecting data for face spoofing detection. Obtaining informed consent from participants and ensuring their privacy and confidentiality are protected throughout the data collection process is essential.

Adhering to ethical guidelines and regulations governing research involving human subjects is also important. In summary, indoor conditions for data collection for smartphone-based spoofing detection involve several key considerations, including lighting, background, camera angles and distances, variety of facial expressions, environmental consistency, participant diversity, and ethical considerations. By following these considerations, researchers can collect reliable and robust data for training and evaluating the spoofing detection system.

3.1.2 Outdoor Acquisition

Collecting data outdoors for smartphone-based spoofing detection presents additional challenges due to the variability of environmental conditions. To effectively navigate these challenges, researchers should consider the following: Firstly, outdoor environments offer diverse lighting conditions, including direct sunlight, shadows, and reflections. Collect data at different times of the day and in various weather conditions to train the spoofing detection system to handle outdoor lighting variability. Secondly, outdoor scenes can include a wide range of backgrounds, such as urban landscapes, parks, or natural settings. Collect data in different outdoor locations to expose the algorithm to various background textures, colors, and patterns. Thirdly, outdoor settings are often dynamic, with moving objects, changing weather conditions, and varying levels of noise. Account for these factors during data collection to simulate real-world scenarios where users may encounter distractions or interruptions while using their smartphones. Fourthly, be prepared to collect data in adverse weather conditions such as rain, fog, or snow. Ensure the smartphone's camera and sensors can function effectively under these conditions, and take appropriate precautions to protect the device from damage. Furthermore, it is crucial to maintain camera stability to capture clear and sharp images despite potential movement or vibration. Consider using stabilization techniques or accessories to minimize motion blur and ensure data quality.

Additionally, respect individuals' privacy and obtain their consent before capturing data in outdoor settings. Inform participants about the purpose of the study and how their data will be used, and provide assurances regarding data security and confidentiality. Finally, annotate outdoor

Fig. 3.2 Outdoor environment setup for experimentation or observation

data with additional contextual information, such as GPS coordinates or timestamps, to facilitate analysis and validation of the spoofing detection algorithm's performance across different outdoor environments. In summary, collecting data outdoors for smartphone-based spoofing detection requires a comprehensive approach that considers the variability of environmental conditions. By taking into account natural lighting, background diversity, dynamic environments, adverse weather conditions, camera stability, privacy and consent, and data annotation, researchers can effectively navigate these challenges and collect reliable and robust data for training and evaluating the spoofing detection system.

The table presents information about various smartphone models, their camera setups, and video recording capabilities. The "Details" column provides additional information about the camera setups, including the type of lenses present and the maximum video recording resolution supported. For smartphones with advanced camera systems like the Samsung S20, iPhone 11Pro, and iPhone 12ProMax, the details include more advanced features like up to 8K or 4K with HDR video recording. For simpler camera setups like those found in the Samsung A03 and Poco C3,

Table 3.1 Smartphone Camera Details: Overview of Camera Setups and Video Recording Capabilities.

Smartphone	Camera	Details
Samsung S20	Rear Camera	Triple-camera setup with wide, ultra-wide, and telephoto lenses
	Video Recording	Capable of recording up to 8K resolution video.
iPhone 11Pro	Rear Camera	Triple-camera setup with wide, ultra-wide, and telephoto lenses.
	Video Recording	Supports 4K video recording.
iPhone 12ProMax	Rear Camera	Triple-camera setup with wide, ultra-wide, and telephoto lenses.
	Video Recording	Capable of recording 4K video with Dolby Vision HDR.
Samsung A03	Rear Camera	Single-camera setup, likely with a standard wide-angle lens.
	Video Recording	Capable of recording video at 720p HD resolution.
Redmi Note 9ProMax	Rear Camera	Quad-camera setup with wide, ultra-wide, macro, and depth sensors.
	Video Recording	Supports 4K video recording.
Poco C3	Rear Camera	Triple-camera setup or single-camera setup.
	Video Recording	Likely supports 1080p Full HD video recording.
Pixel 7	Rear Camera	Likely dual or triple-camera setup.
	Video Recording	Capable of recording 4K video.
Galaxy S8	Rear Camera	Single camera.
	Video Recording	Capable of recording 4K video.
iPhone 8	Rear Camera	Single camera.
	Video Recording	Supports 4K video recording.
Galaxy S10	Rear Camera	Triple-camera setup with wide, ultra-wide, and telephoto lenses.
	Video Recording	Capable of recording 4K video.

the details focus on whether the camera setup is single or triple, and the maximum resolution supported for video recording. The table also provides information about additional camera features such as quad-camera setups and their video recording capabilities.

Table 3.2 Smartphone Camera Details: Overview of Camera Setups and Video Recording Capabilities table 2

Smartphone	Camera	Details
Galaxy S10	Rear Camera	Triple-camera setup with wide, ultra-wide, and telephoto lenses.
	Video Recording	Capable of recording 4K video.
Samsung Galaxy J7	Rear Camera	Single-camera setup, likely with a standard wide-angle lens.
	Video Recording	Capable of recording video at 1080p Full HD resolution
samsung S20	Rear Camera	Triple-camera setup with wide, ultra-wide, and telephoto lenses.
	Video Recording	Capable of recording up to 8K resolution video.
Redmi Note 9ProMax	Rear Camera	Quad-camera setup with wide, ultra-wide, macro, and depth sensors.
	Video Recording	Supports 4K video recording.

A table compares the camera setups and video recording capabilities of four smartphones: Galaxy S10, Samsung Galaxy J7, Samsung S20, and Redmi Note 9ProMax. The table includes details about the rear camera setup and video recording capabilities for each smartphone. The "Camera" column specifies the type of camera setup and the lenses present, while the "Details" column provides information about the maximum resolution supported and any additional features such as HDR support. The table is formatted with vertical and horizontal lines, with centered content within each cell and resized to fit within the text width of the document. Overall, the table provides a comparison of the camera features of different smartphones.

The data collection for evaluating the performance of smartphone models in detecting face mask attacks involved diverse participants wearing different types of masks, including silicon mask ,letax mask,paper,joker and so on. The experiments were conducted in various environments, such as indoor, outdoor to check the effectiveness of the smartphone models. The camera positions were varied to capture the facial features accurately, and consistent lighting and varied backgrounds were maintained in the indoor experiments. Adverse weather conditions were simulated in the outdoor experiments to evaluate the performance of the smartphone models in detecting face mask attacks in real-world scenarios.

Table 3.3 Data Collection Protocol for Facial Data in Surveillance Setups: Overview of Categories, Subjects, Protocols, and Conditions.

Category	Data	Subject	Protocols	Smart phone	2 Variants (Glass & NoGlass)	Sessions	Conditions (indoor & outdoor)	Number of sample at 30fps
Bonafide (BF)	BF	30	1	10	2	1	2	10
Attack	SMA	4	2	10	2	1	2	30
	LMA	3	2	10	2	1	2	30
	WMA	10	1	10	–	1	2	15
	PPA	10	1	10	–	1	2	20
	PHMA	4	2	10	2	1	2	20
	JCPHMA	3	2	10	2	1	2	20
	CPHMA	3	2	10	2	1	2	20

The table presents information on different categories of data, including "BF" for Bonafide data and various types of "Attack" scenarios involving spoofing attacks. The data is represented by abbreviations or codes for each category, while the subject column denotes the number of individuals involved in data collection. The protocols column shows the number of specific procedures followed during data collection, and the smartphone column specifies the number of smartphones used. The table also indicates the presence of multiple variants of attack scenarios, such as with or without glasses, and the number of data collection sessions conducted for each category. Additionally, the table distinguishes between indoor and outdoor environmental conditions under which data collection occurred and provides the total number of samples collected at a frame rate of 30 frames per second for each category.

Table 3.4 Abbreviations and Their Full Forms: Reference Table.

SHORT FORM	**FULL FORM**
BF	Bonafide
SMA	Silicon mask attack
LMA	Latex Mask Attack
WMA	Wrap mask attack
PHMA	Plastic Hard Mask Attack
JCPHMA	Joker Color Plastic Hard Mask Attack
CPHMA	Color Plastic Hard Mask Attack
PMA	Paper Mask Attack
EMA	Eye Mask Attack

The table includes two columns, one listing the abbreviations or short forms used and the other providing their corresponding full forms or expanded versions. The abbreviations and their full forms listed in the table are as follows: "BF" stands for Bonafide, "SMA" for Silicon Mask Attack, "LMA" for Latex Mask Attack, "WMA" for Wrap Mask Attack, "PHMA" for Plastic Hard Mask Attack, "JCPHMA" for Joker Color Plastic Hard Mask Attack, "CPHMA" for Color Plastic Hard Mask Attack, "PMA" for Paper Mask Attack, and "EMA" for Eye Mask Attack.

3.2 Bonafide

In the context of face recognition, the term "bonafide data" refers to authentic facial images that are obtained from real individuals. These images are usually acquired under carefully controlled conditions to ensure that they faithfully represent the distinct facial features of each person.

3.3 Presentation Attach Instruments (PAI)

3.3.1 Silicon Face Mask Attack

Silicon mask attack, also known as a mask presentation attack, refers to a class of spoofing or presentation attacks that are designed to circumvent facial recognition systems. This type of attack is executed by using a highly realistic mask made of silicon or other materials to impersonate a genuine user. Its primary objective is to deceive the facial recognition system into recognizing the attacker as the authorized user, thereby providing access to confidential systems or sensitive information.

3.3.2 Latex Mask Attack

In a Latex Mask Attack, an attacker utilizes a mask made of latex or similar material to impersonate a genuine user. These masks are crafted to resemble human faces and can be designed with intricate details to closely mimic facial features. The attacker presents the latex mask to the facial recognition system, aiming to bypass authentication and gain unauthorized access.

Fig. 3.3 Illustrations of Silicon Mask Presentation Attacks: Visual Examples of Facial Impersonation Using Silicon Masks.

Fig. 3.4 Latex Mask Examples: Depictions of Facial Presentation Attacks Utilizing Latex Masks.

3.3.3 Wrap Mask Attack

A Wrap Mask Attack involves the use of a thin, flexible material wrapped around the face to create a 3D representation of a genuine user's face. The material used for wrapping can vary, including paper, fabric, or even specially designed materials. Despite being less elaborate than latex masks, wrap masks can still deceive facial recognition systems by creating the illusion of facial depth and contours.

3.3.4 Plastic soft Mask Attack

In a Plastic soft Mask Attack, an attacker employs a rigid mask made of plastic or other hard materials to impersonate a genuine user. Unlike latex masks, which are flexible, plastic hard masks maintain their shape and structure more effectively. These masks may be 3D-printed or manually crafted to closely resemble human faces, making them difficult for facial recognition systems to differentiate from genuine faces.

Fig. 3.5 Illustrations of Plastic Soft Mask Presentation Attacks: Visual Depictions of Facial Impersonation Using Soft Plastic Masks.

Fig. 3.6 Joker Color Plastic Hard Mask Attacks: Depictions of Facial Impersonation Using Vibrant Hard Masks.

3.3.5 Joker Color Plastic Hard Mask Attack

This attack involves the use of a rigid plastic mask, similar to the Plastic Hard Mask Attack, but with additional coloring or design elements inspired by the Joker character from popular culture. The purpose is to create a visually striking mask that may distract or deceive facial recognition systems, making it difficult for them to differentiate between the mask and a genuine human face.As show in the fig 3.6.

3.3.6 Color Plastic Hard Mask Attack

In a Color Plastic Hard Mask Attack, the attacker utilizes a rigid plastic mask that is colored or painted to resemble a genuine human face. These masks may incorporate skin tones, facial features, and other details to create a realistic appearance. By presenting the color plastic hard mask to the facial recognition system, the attacker aims to bypass authentication and gain unauthorized access.As show in fig.3.7.

3.3.7 Paper Mask Attack

A Paper Mask Attack involves the use of a mask made of paper or similar lightweight materials. While less durable and realistic compared to plastic or latex masks, paper masks can still be used to deceive facial recognition systems, especially in scenarios where the system's security

Fig. 3.7 Color Plastic Hard Mask Presentation Attacks: Visual Examples of Facial Impersonation Using Hard Masks in Various Colors

measures are minimal or easily bypassed. Attackers may create paper masks with printed facial features or hand-drawn designs to mimic a genuine human face.

Fig. 3.8 Paper Mask Presentation Attacks: Depictions of Facial Impersonation Using Paper Masks.

3.3.8 Eye mask attack

In an Eye Mask Attack, the attacker focuses specifically on creating a mask that covers only the eyes while leaving the rest of the face exposed. By strategically concealing key facial features such as the nose, mouth, and chin, the attacker aims to exploit vulnerabilities in the facial recognition system's detection algorithms, potentially tricking the system into granting unauthorized access.

Fig. 3.9 Eye Mask Presentation Attacks: Visual Examples of Facial Impersonation Using Eye Masks.

3.4 Data Acqusition setup

The data capture process involved two environments, indoor and outdoor. The indoor environment was a corridor with approximately 9-15 incandescent bulbs that were 200 watts each and used as lighting. On the other hand, the outdoor data collection was done inside a walking passage. During data capture, the distance between the sensor and the subject was measured in two parts: the first distance was from the subject's start to stop, which was approximately 8.5 meters, and the second distance was from the subject's stop to the sensor, which was approximately 1.4 meters.

Below is a list of different scenarios where facial recognition systems can be attacked, along with their descriptions:

- **Bonafide (BF):** Data collection occurs for Glass and No Glass conditions in a single session, with participants wearing a Hoodie and Cap.
- **Latex Mask Attack (LMA):** Data capture involves two protocols within a single session. In Protocol 1, participants don a Hoodie, Cap, and Latex Mask, while in Protocol 2, they wear a Wig and Latex Mask.

- **Silicon Mask Attack (SMA):** Similar to LMA, this attack comprises two protocols in one session. Protocol 1 involves participants wearing a Hoodie, Cap, and Silicon Mask, while Protocol 2 involves a Wig, Cap, and Silicon Mask.
- **Wrap Mask Attack (WMA):** Data is collected in one protocol during a session where participants wear a color-printed face photo wrap.
- **Printed Photo Attack (PPA):** Participants use a color-printed face photo in a single protocol session for data capture.
- **Plastic Hard Mask Attack (PHMA):** Data collection involves two protocols during one session. Protocol 1 includes participants wearing a Hoodie, Cap, and Plastic Hard Mask, while Protocol 2 involves a Wig, Cap, and Plastic Hard Mask.
- **Joker Color Plastic Hard Mask Attack (JPHMA):** Similar to PHMA, this attack comprises two protocols in one session. Protocol 1 involves participants wearing a Hoodie, Cap, and Joker Color Plastic Hard Mask, while Protocol 2 involves a Wig, Cap, and Joker Color Plastic Hard Mask.
- **Color Plastic Hard Mask Attack (CPHMA):** Participants undergo data collection for Glass and No Glass in two protocols during a single session. Protocol 1 involves wearing a Hoodie, Cap, and Color Plastic Hard Mask, while Protocol 2 includes a Wig, Cap, and Color Plastic Hard Mask.
- **Paper Mask Attack (PMA):** Two protocols are executed during a single session, with participants wearing a Hoodie, Cap, and Paper Mask in Protocol 1, and a Wig, Cap, and Paper Mask in Protocol 2.
- **Eye Mask Attack (EMA):** Data capture occurs in one protocol session where participants wear a Halloween color eye mask.

This table provides a summary of the total time required for data collection across various scenarios involving different types of facial masks. The data includes the number of subjects,

Table 3.5 Total Time Required for Data Collection

Data	subject	variants	time	sec	Total time
Bonafide	25	2	15	30	
Silicon mask	4	4	30	30	375
Latex Mask	4	4	30	30	240
Wrap mask	10	1	10	30	240
Print Photo mask	10	1	10	30	50
Plastic Hard Mask	4	2	30	30	50
Plastic Soft Mask	4	2	30	30	120
Color Plastic Hard Mask	3	2	30	30	90
Paper Mask	2	1	30	30	30
Eye Mask	5	1	30	30	75

variants, time taken per session (in seconds), time per subject (in seconds), and the total time taken for data collection (in seconds) for each scenario.

- **Bonafide:** Data collection involving 25 subjects, with 2 variants, each taking 15 seconds, resulting in a total time of 30 seconds.
- **Silicon Mask:** Data collection for silicon mask scenarios, involving 4 subjects and 4 variants, each taking 30 seconds, resulting in a total time of 375 seconds.
- **Latex Mask:** Similar to the silicon mask scenario, data collection involves 4 subjects and 4 variants, each taking 30 seconds, resulting in a total time of 240 seconds.
- **Wrap Mask:** Data collection for the wrap mask scenario involves 10 subjects, with 1 variant, each taking 10 seconds, resulting in a total time of 240 seconds.
- **Print Photo Mask:** Data collection for the print photo mask scenario involves 10 subjects, with 1 variant, each taking 10 seconds, resulting in a total time of 50 seconds.
- **Plastic Hard Mask:** Data collection for the plastic hard mask scenario involves 4 subjects, with 2 variants, each taking 30 seconds, resulting in a total time of 50 seconds.

- **Plastic Soft Mask:** Similar to the plastic hard mask scenario, data collection involves 4 subjects and 2 variants, each taking 30 seconds, resulting in a total time of 120 seconds.
- **Color Plastic Hard Mask:** Data collection for the color plastic hard mask scenario involves 3 subjects, with 2 variants, each taking 30 seconds, resulting in a total time of 90 seconds.
- **Paper Mask:** Data collection for the paper mask scenario involves 2 subjects, with 1 variant, each taking 30 seconds, resulting in a total time of 30 seconds.
- **Eye Mask:** Data collection for the eye mask scenario involves 5 subjects, with 1 variant, each taking 30 seconds, resulting in a total time of 75 seconds.

Chapter 4

Methodology

In the realm of presentation attack detection (PAD), feature-based methods are crucial for identifying spoof attempts by processing various features extracted from facial images. These features encompass texture, temporal data, image quality, and vital signs. These methods can be broadly categorized into static and dynamic approaches.

Static approaches, exemplified by texture and image quality analysis, operate without relying on temporal information. They assess each frame individually, allowing for video-based anti-spoofing tasks. The culmination of assessments from multiple frames informs the final decision, making static methods popular due to their efficiency, low computational requirements, and cost-effectiveness. In contrast, dynamic approaches leverage temporal information, analyzing motion or life signs to verify liveness in facial recognition (FR) systems. While dynamic methods offer enhanced spoof detection capabilities, they typically require more computational resources.

Texture-based PAD methods excel in differentiating genuine images from fakes through micro-textural analysis, particularly in detecting photo and replay attacks. The utilization of Local Binary Pattern (LBP) descriptors is prevalent in these methods, although they may struggle with low-resolution images. Presentation attacks often introduce image distortions such as surface reflections, color distortions, and shape deformations, which are exploited by face PAD systems.

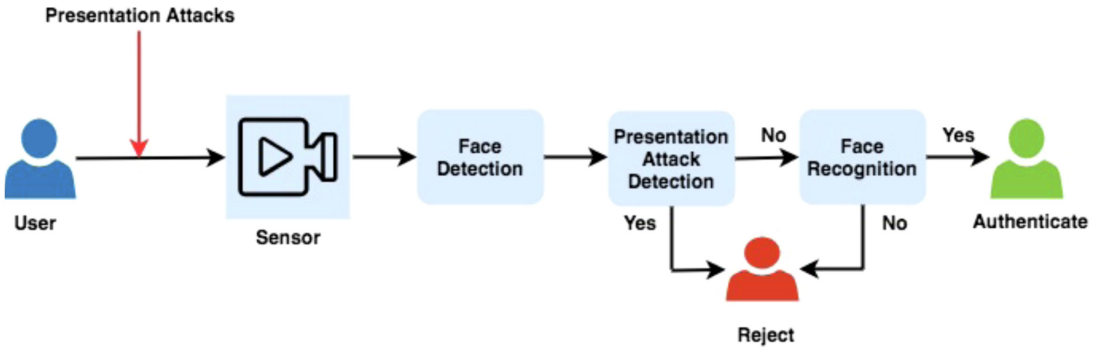


Fig. 4.1 Overview of Presentation Attack Detection Techniques in Biometric Systems

Dynamic approaches rely on temporal feature analysis, considering relative motion in videos for spoof detection. Some methods utilize life signs such as pulse, eye blinking, or lip movement to confirm liveness. Techniques like Dynamic Mode Decomposition (DMD) utilize temporal cues like eye blinking and lip movements for liveness identification. However, motion-based techniques may demand user cooperation, impacting processing time.

Traditional feature-based methods predominantly relied on hand-crafted features like Local Binary Patterns (LBP), Histogram of Oriented Gradient descriptors (HOG), etc., which often struggled with generalization due to variations in spoofing mediums and devices. The advent of deep learning has revolutionized feature learning, leading to superior detection performance compared to hand-crafted methods. Consequently, there's been a significant shift towards deep learning-based approaches in face PAD.

The enrollment process involves capturing a video of the individual's face using the smartphone camera, ensuring that the face is well-illuminated and clearly visible. The captured video is then split into individual frames, and a face detection algorithm is applied to each frame to detect and locate the person's face. Subsequently, facial features are extracted from the detected face regions, which may include landmarks, texture descriptors, or deep learning embeddings that represent unique characteristics of the person's face. The extracted facial features are utilized as input to train a classification model (such as SVM or CNN), and labels are assigned to indicate whether each sample is bonafide or a mask attack. By learning to distinguish between genuine faces and mask attacks based on the extracted features, the model effectively classifies new

incoming samples as either bonafide or fraudulent.

To capture a new sample, a video of the person's face, either bonafide or wearing a mask, is captured and split into individual frames. A face detection algorithm is applied to each frame to detect and locate the person's face. Subsequently, the trained classification model is applied to the extracted facial features from each frame to predict whether each frame corresponds to a bonafide face or a mask attack. To evaluate the accuracy of the classifier, the predictions are compared to manually annotated ground truth labels for the sample video based on whether the person is genuinely presenting their face or wearing a mask during the sample capture. Based on the classifier's predictions and the accuracy evaluation, it can be determined whether each sample video corresponds to a bonafide face or a mask attack.

4.1 Block Diagram:

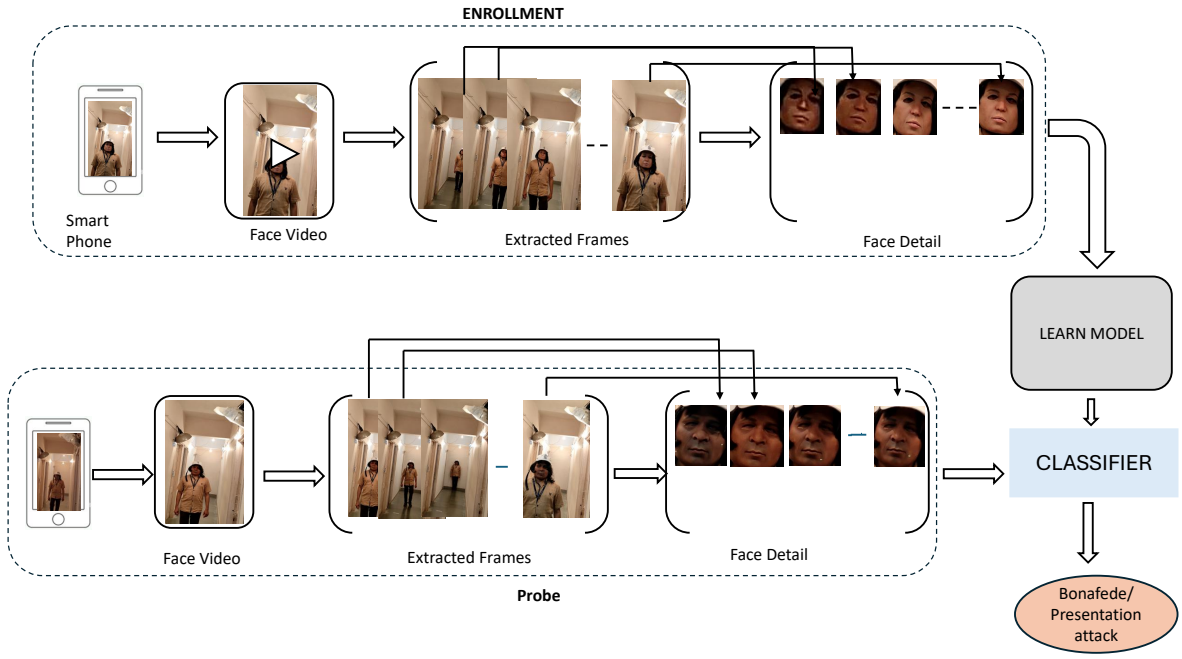


Fig. 4.2 Block Diagram Illustrating Presentation Attack Detection Methodology in Facial Recognition Systems

4.1.1 Data recording using smartphone

To capture a video using a smartphone, there are several general steps that can be followed. Firstly, the camera app on the smartphone should be opened. This can usually be done from the home screen or the app drawer. Next, the user should switch to video mode, which is typically accessible via an icon or option labeled "Video" or a camcorder icon. Once the user has entered video mode, they should frame their shot using the smartphone's screen, adjusting the camera angle, zoom, and focus as desired. Recording can then be initiated by tapping the "Record" button, typically represented by a circle icon. The recording indicator will appear on the screen, indicating that the video is being recorded. The user can capture the desired footage by keeping the camera steady and recording. Recording can be stopped at any time by tapping the "Stop" button, usually represented by a square icon. After recording, the user can review the captured video in the camera app, where they may have the option to preview the video, trim the footage, or apply filters or effects before saving it to their smartphone's gallery.

4.1.2 Frame Extraction

The process of extracting flames from videos involves several preprocessing steps that are designed to isolate the regions of interest. Firstly, the video needs to be split into individual frames, with each frame representing a single image in the video sequence. Next, the frames are converted to a color space that enhances flame visibility, such as the HSV (Hue, Saturation, Value) color space, where flames typically have distinct hue and intensity characteristics. After this, image processing techniques are applied to detect potential flame regions in each frame, such as thresholding, edge detection, or blob analysis. Once these regions are identified, they are segmented from the background using techniques like connected component analysis or region growing to isolate the flames more accurately. The detected flame regions are then subjected to filtering or refinement techniques to remove noise and false positives, such as morphological operations, spatial filtering, or machine learning-based classification. Optionally, additional post-processing steps can be performed to enhance the quality or clarity of the extracted flame regions

using techniques like smoothing, sharpening, or contrast adjustment. Finally, the extracted flame regions are saved or displayed for further analysis or visualization. These preprocessing steps can be implemented using various image processing libraries and tools such as OpenCV, MATLAB, or Python with libraries like NumPy and SciPy, with specific techniques and parameters being selected based on factors such as the characteristics of the input video, the complexity of the flame scenes, and the desired level of accuracy in flame extraction.

4.1.3 Frame rate

The number of frames captured by a smartphone camera in a 30-second video can be calculated based on the frame rate (fps) of the video recording and camera resolution settings. The frame rate refers to the number of frames captured or displayed per second, and typical frame rates for smartphone video recording include 24 fps, 30 fps, 60 fps, and higher. To calculate the number of frames in a 10-second video, multiply the frame rate by the duration of the video (in seconds). For example, if the frame rate is 30 fps, the number of frames in a 10-second video would be 30 x 30 = 900 frames. However, it's important to note that the actual number of frames captured can vary depending on the camera's resolution settings, as higher resolutions may require more processing power and storage capacity, which can affect the frame rate and number of frames captured.

To calculate the total number of frames

Total number of frames=Frame rate (fps)×Duration (seconds)

4.1.4 Face Detection

Detecting faces in video frames involves a series of steps, which include frame extraction, preprocessing, face detection, face region extraction, and optional post-processing. Firstly, the video should be split into individual frames, and each frame is preprocessed to enhance features

relevant to face detection. This can include resizing, converting to grayscale, and histogram equalization to improve contrast. Next, a face detection algorithm is applied to each preprocessed frame to detect and locate faces. There are several popular algorithms for face detection, such as Haar Cascade Classifiers, Histogram of Oriented Gradients (HOG), and Convolutional Neural Networks (CNNs). Once faces are detected, the bounding boxes or regions of interest containing the detected faces are extracted from the frames. Post-processing techniques can be applied to refine the detected face regions, such as non-maximum suppression to remove redundant detections or smoothing to stabilize the bounding boxes across frames. Finally, the frames with the detected face regions highlighted can be visualized or saved, or the face regions can be extracted for further analysis or processing.

4.1.5 Learning model

In the realm of machine learning, a learning model is a mathematical representation of a problem domain that a machine learning algorithm uses to make predictions or decisions based on input data. The models are trained on labeled data to identify patterns and relationships within the data, enabling them to make predictions on new, unseen data. There are different types of learning models, including regression models, classification models, clustering models, dimensionality reduction models, and deep learning models. Regression models are used to predict continuous numerical values, while classification models categorize input data into discrete classes or categories. Clustering models group similar data points together based on their features, and dimensionality reduction models reduce the number of features in a dataset while preserving important information. Deep learning models are a subset of neural network models with multiple layers of interconnected neurons. The selection of a learning model depends on the nature of the problem, the characteristics of the data, and the desired outcome. The effectiveness of different models varies depending on the specific task and dataset. Therefore, choosing the appropriate learning model is crucial in designing effective machine learning systems.

Presentation attack detection (PAD) is a vital aspect of face recognition systems, where the objective is to identify and distinguish between genuine and fake facial presentations. Various algorithms and techniques can be used for PAD, including traditional computer vision techniques, machine learning models, feature fusion, liveness detection, deep learning architectures, and adversarial training. Traditional computer vision techniques such as texture analysis, feature extraction, and motion analysis can be utilized to detect anomalies in facial images and differentiate between genuine and fake faces. SVM classifiers, CNN models, and Pro CRC can be trained on features extracted from facial images to identify presentation attacks and distinguish them from genuine facial presentations. Feature fusion techniques, such as combining texture, motion, and shape features, can enhance the robustness of the detection system. Liveness detection methods, such as analyzing physiological signals or requiring specific actions, can verify the presence of live human subjects during facial recognition. Advanced deep learning architectures tailored for PAD tasks, such as Siamese networks, Capsule networks, or attention mechanisms, can capture complex patterns and variations in facial images and improve the accuracy of presentation attack detection. Adversarial training techniques can also be used to improve the robustness of the model against adversarial attacks aimed at bypassing the detection system. In the context of the algorithms mentioned, SVM, Pro CRC, and AlexNet can be utilized for PAD tasks to distinguish between genuine and spoofed faces. By combining these algorithms and techniques, presentation attack detection systems can effectively identify and mitigate the risks associated with spoofing attacks in facial recognition systems, ensuring the reliability and security of biometric authentication processes.

4.2 Unleashing the Power of Convolutional Neural Networks (CNNs) in Machine Learning: A Revolution in Computer Vision and Beyond

Convolutional Neural Networks (CNNs) have marked a significant breakthrough in the realm of machine learning, particularly in the domain of computer vision. These deep learning models are tailored to effectively process and analyze structured grid-like data, with images as their primary input. Unlike traditional neural networks, CNNs feature a hierarchical architecture comprising specialized layers that extract hierarchical representations of visual data. The key components of CNNs are the convolutional layers, which utilize learnable filters or kernels to capture local patterns and features across the input image. These convolutional operations are complemented by pooling layers, which downsample the spatial dimensions of the feature maps, enabling hierarchical feature extraction while reducing computational complexity. Through supervised learning on large datasets, CNNs autonomously learn to extract hierarchical features from raw image data, progressively discerning complex visual patterns and structures. Additionally, CNN architectures often incorporate fully connected layers at the end of the network to synthesize the extracted features for producing the final output, such as class probabilities in image classification tasks.

Beyond their application in image classification, CNNs have been extended to a wide range of computer vision tasks, including object detection, semantic segmentation, image generation, and more. Moreover, CNNs have transcended the domain of computer vision and found utility in diverse fields like natural language processing, speech recognition, and reinforcement learning. The versatility and effectiveness of CNNs have solidified their position as a cornerstone of contemporary machine learning, empowering researchers and practitioners to address complex real-world problems with unprecedented accuracy and efficiency. As CNNs continue to evolve, researchers are exploring novel architectures, training techniques, and applications to further extend their capabilities and push the boundaries of what is possible with deep learning.

The impact of CNNs goes beyond traditional computer vision tasks, permeating various other domains of machine learning and artificial intelligence. CNNs have been applied to tasks such as text classification, sentiment analysis, and language translation in natural language processing, where input data is represented as sequences of words or characters. Similarly, in speech recognition, CNNs have been employed to process audio signal spectrograms, enabling accurate transcription and understanding of spoken language.

As CNNs continue to evolve, researchers are exploring novel architectures, training techniques, and applications to further push the boundaries of what is possible with deep learning. Recent advancements include attention mechanisms, capsule networks, and self-supervised learning, aiming to improve model interpretability, robustness, and efficiency. However, the ethical and societal implications of CNNs are increasingly being scrutinized, with efforts underway to address issues like bias, fairness, privacy, and accountability in the deployment of AI systems.

In conclusion, CNNs have transformed the landscape of machine learning, offering unparalleled capabilities for processing and understanding visual data. Their versatility, scalability, and effectiveness have catalyzed transformative advancements across a wide range of fields, driving progress and innovation in the era of artificial intelligence. As CNNs continue to evolve, it is essential to ensure that their deployment is ethical, fair, and transparent, benefiting society as a whole.

4.2.1 Alexnet

AlexNet is a convolutional neural network (CNN) architecture that was introduced in 2012 by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton. It is one of the pioneering deep learning models that has played a significant role in advancing computer vision tasks, particularly image classification. The architecture of AlexNet comprises eight layers of learnable parameters, including five convolutional layers followed by max-pooling layers and three fully connected layers. Dropout layers are also incorporated to prevent overfitting. AlexNet's convolutional layers use a kernel size of 3x3 with a stride of 1 and rectified linear unit (ReLU) activation

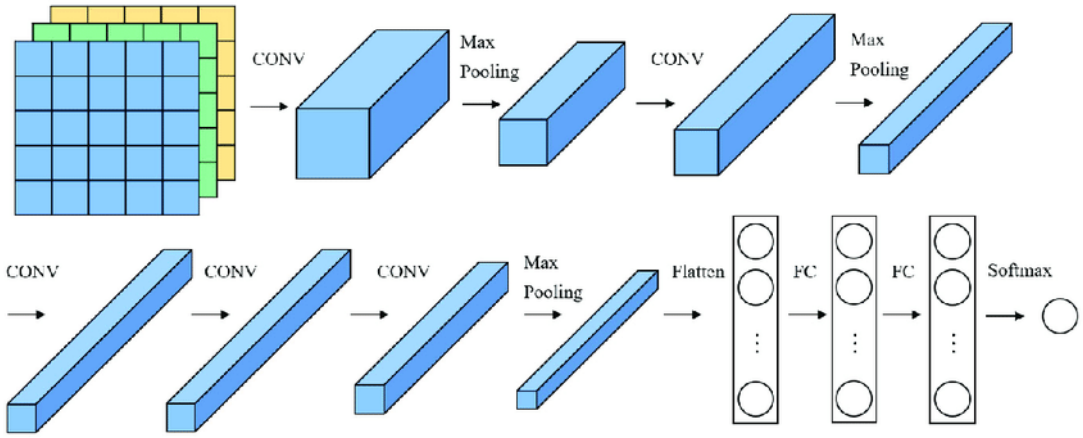


Fig. 4.3 AlexNet: A Pioneering Deep Convolutional Neural Network Architecture for Image Classification

functions. These layers are responsible for extracting features from the input images. Additionally, max-pooling layers are applied after each set of convolutional layers to reduce the spatial dimensions of the feature maps and introduce translation invariance. ReLU activation functions are used throughout the network, which helps in mitigating the vanishing gradient problem and accelerating the training process. Local Response Normalization (LRN) is applied after the first and second convolutional layers in AlexNet. This technique normalizes the activity of neurons across adjacent channels, enhancing the model's generalization ability. The last three layers of AlexNet are fully connected layers that are responsible for classification based on the features extracted by the preceding convolutional layers. The final layer employs a softmax activation function to output class probabilities. AlexNet was trained on the ImageNet dataset, which contains millions of labeled images across thousands of classes. The training process involved using stochastic gradient descent (SGD) with momentum, data augmentation techniques, and dropout regularization. The success of AlexNet on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012 marked a significant breakthrough in the field of deep learning.

AlexNet works by employing a deep convolutional neural network (CNN) architecture to process and classify images. Here's a simplified overview of how it operates:

Input Images: AlexNet takes images as input. These images are typically represented as matrices of pixel values, where each pixel corresponds to a color value (e.g., red, green, blue). Convolutional Layers: The input image is passed through a series of convolutional layers. Each convolutional layer consists of a set of learnable filters (also called kernels) that slide across the input image. These filters convolve with the input image, extracting features such as edges, textures, and patterns. The convolution operation is performed by taking the dot product between the filter and a small region of the input image, producing a feature map. Activation Function (ReLU): After each convolutional operation, a rectified linear unit (ReLU) activation function is applied element-wise to the feature map. ReLU introduces non-linearity into the network and helps in capturing complex patterns in the data. Pooling Layers: Following some of the convolutional layers, max-pooling layers are applied. Max-pooling reduces the spatial dimensions of the feature maps by selecting the maximum value within each region of the feature map. This downsampling helps in reducing computational complexity and makes the network more robust to variations in input. Local Response Normalization (LRN): In AlexNet, local response normalization (LRN) is applied after the first and second convolutional layers. LRN normalizes the activation of neurons across adjacent channels, enhancing the network's ability to generalize. Fully Connected Layers: After several convolutional and pooling layers, the feature maps are flattened into a vector and fed into fully connected layers. These layers perform high-level reasoning and classification based on the features extracted by the convolutional layers. The output of the last fully connected layer is typically passed through a softmax activation function to produce class probabilities.[6]

Training: AlexNet is trained using supervised learning with labeled data. During training, the model's parameters (weights and biases) are adjusted iteratively using optimization algorithms such as stochastic gradient descent (SGD) with momentum. The goal is to minimize a loss function that quantifies the difference between the predicted outputs and the ground truth labels. Data augmentation techniques, dropout regularization, and LRN are used to prevent overfitting and improve generalization. Output: The final output of AlexNet is a probability distribution over the classes in the dataset. It predicts the probability that the input image belongs to each

class, allowing for tasks such as image classification.

4.3 Exploring Traditional Methods in Face Recognition: Leveraging Classical Machine Learning and Computer Vision Techniques

Before the emergence of deep learning and neural network-based approaches, traditional methods in the context of face recognition were prevalent. These methods relied on classical machine learning algorithms and computer vision techniques. Here are some key points about traditional methods in face recognition:

Feature-Based Approaches: Traditional methods often involved the use of handcrafted features extracted from facial images. These features included geometric features such as distances between facial landmarks, texture features like local binary patterns (LBP), and appearance-based features like Eigenfaces or Fisherfaces.

Dimensionality Reduction: Many traditional methods incorporated dimensionality reduction techniques to simplify the feature space and enhance computational efficiency. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) were commonly used for this purpose.Classification Algorithms: Traditional classification algorithms such as Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), and Decision Trees were frequently employed for face recognition tasks. These algorithms learned a mapping from the extracted features to class labels and could distinguish between different individuals based on their facial characteristics. Template Matching: Some traditional methods utilized template matching techniques to compare facial images. These techniques involved comparing a test image with a database of reference images to find the closest match based on predefined similarity measures.

Challenges: Traditional methods often faced challenges in handling variations in lighting conditions, facial expressions, occlusions, and pose changes. They also struggled with managing large datasets and required manual parameter tuning.

Robustness and Efficiency: Despite their limitations, traditional methods were lauded for their simplicity, interpretability, and computational efficiency. They performed well in constrained environments with controlled conditions and limited variability.

Integration with Deep Learning: In recent years, there has been a trend towards integrating traditional methods with deep learning techniques to leverage the strengths of both approaches. For instance, deep learning models may be used for feature extraction, while traditional classifiers are employed for classification tasks.

4.3.1 support vector machines (SVM)

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:[14]

Example: SVM can be understood with the example that we have used in the KNN classifier. Suppose we see a strange cat that also has some features of dogs, so if we want a model that can accurately identify whether it is a cat or dog, so such a model can be created by using the SVM algorithm. We will first train our model with lots of images of cats and dogs so that it can learn about different features of cats and dogs, and then we test it with this strange creature. So as support vector creates a decision boundary between these two data (cat and dog) and choose

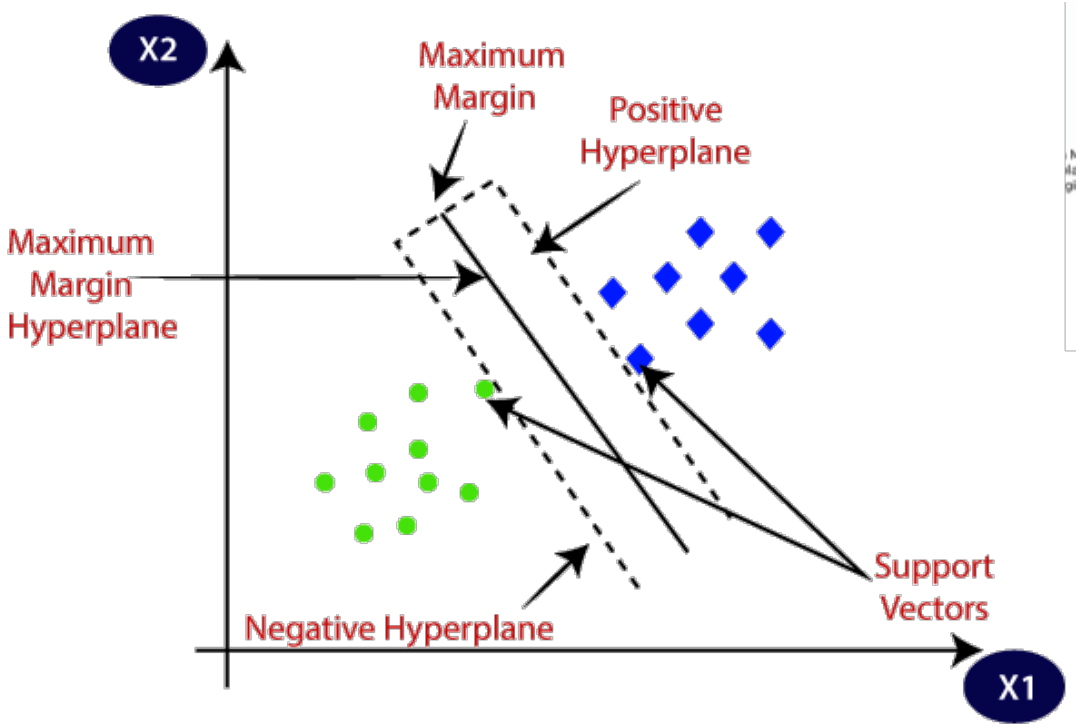


Fig. 4.4 Visualization of Support Vector Machine (SVM) Decision Boundaries for Classification

extreme cases (support vectors), it will see the extreme case of cat and dog. On the basis of the support vectors, it will classify it as a cat. Consider the below diagram:

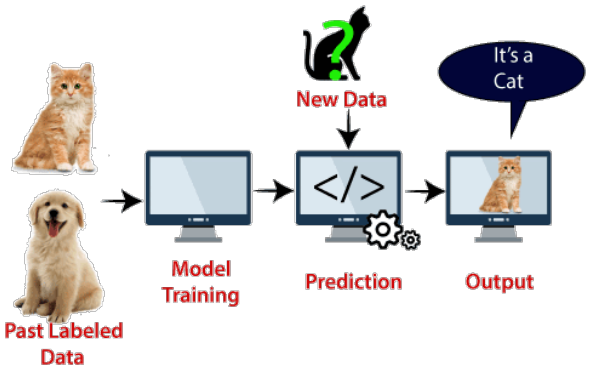


Fig. 4.5 SVM algorithm can be used for Face detection, image classification, text categorization, and so on

[14]

Types of SVM classifiers

Linear SVM are used with linearly separable data; this means that the data do not need to undergo any transformations to separate the data into different classes. The decision boundary

and support vectors form the appearance of a street, and Professor Patrick Winston from MIT uses the analogy of "fitting the widest possible street"2 (link resides outside ibm.com) to describe this quadratic optimization problem. Mathematically, this separating hyperplane can be represented as:

wx + b = 0

where w is the weight vector, x is the input vector, and b is the bias term.

There are two approaches to calculating the margin, or the maximum distance between classes, which are hard-margin classification and soft-margin classification. If we use a hard-margin SVMs, the data points will be perfectly separated outside of the support vectors, or "off the street" to continue with Professor Hinton's analogy. This is represented with the formula,

$$(wx_j + b)y_j \geq a \tag{4.1}$$

and then the margin is maximized, which is represented as: $\max Y = \frac{a}{\|W\|}$ where a is the margin projected onto W. Soft-margin classification is more flexible, allowing for some misclassification through the use of slack variables ξ The hyperparameter, C, adjusts the margin; a larger C value narrows the margin for minimal misclassification while a smaller C value widens it, allowing for more misclassified data.

Nonlinear SVM Much of the data in real-world scenarios are not linearly separable, and that's where nonlinear SVMs come into play. In order to make the data linearly separable, preprocessing methods are applied to the training data to transform it into a higher-dimensional feature space. That said, higher dimensional spaces can create more complexity by increasing the risk of overfitting the data and by becoming computationally taxing. The "kernel trick" helps to reduce some of that complexity, making the computation more efficient, and it does this by replacing dot product calculations with an equivalent kernel function4.

There are a number of different kernel types that can be applied to classify data. Some popular kernel functions include:

- Polynomial kernel
- Radial basis function kernel (also known as a Gaussian or RBF kernel)
- Sigmoid kernel

Support vector regression (SVR) Support vector regression (SVR) is an extension of SVMs, which is applied to regression problems (i.e. the outcome is continuous). Similar to linear SVMs, SVR finds a hyperplane with the maximum margin between data points, and it is typically used for time series prediction. SVR differs from linear regression in that you need to specify the relationship that you're looking to understand between the independent and dependent variables. An understanding of the relationships between variables and their directions is valuable when using linear regression. This is unnecessary for SVRs as they determine these relationships on their own. TutorialClassifying data using the SVM algorithm using Python Use SVMs with scikit-learn to make predictions accounts likely to default on their credit card.

Hyperplane and Support Vectors in the SVM algorithm

Hyperplane: There can be multiple lines/decision boundaries to segregate the classes in n-dimensional space, but we need to find out the best decision boundary that helps to classify the data points. This best boundary is known as the hyperplane of SVM.

The dimensions of the hyperplane depend on the features present in the dataset, which means if there are 2 features (as shown in image), then hyperplane will be a straight line. And if there are 3 features, then hyperplane will be a 2-dimension plane.

We always create a hyperplane that has a maximum margin, which means the maximum distance between the data points.

Support Vectors

The data points or vectors that are the closest to the hyperplane and which affect the position of the hyperplane are termed as Support Vector. Since these vectors support the hyperplane, hence called a Support vector.

How SVMs work

The working of the SVM algorithm can be understood by using an example. Suppose we have a dataset that has two tags (green and blue), and the dataset has two features x1 and x2. We want a classifier that can classify the pair(x1, x2) of coordinates in either green or blue. Consider the below image:

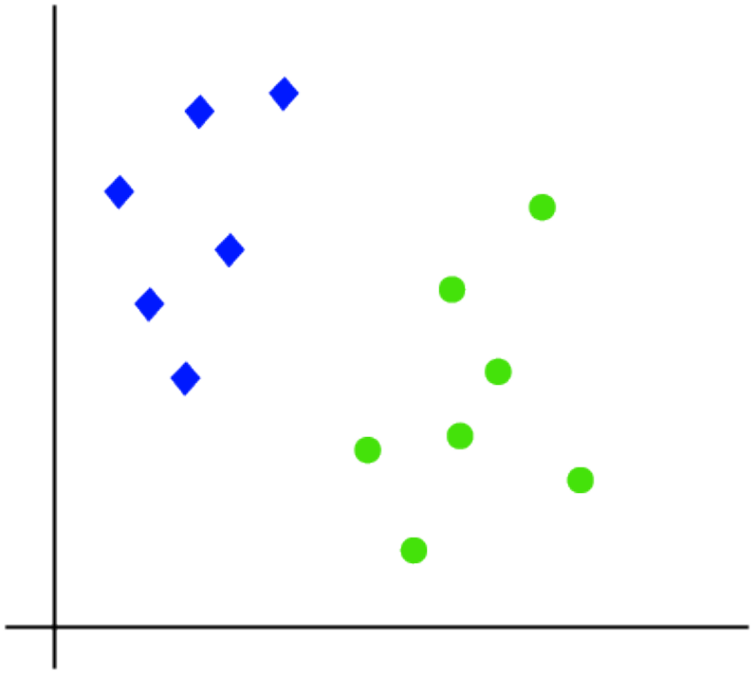


Fig. 4.6 Example Illustrating SVM Algorithm for Binary Classification

So as it is 2-d space so by just using a straight line, we can easily separate these two classes. But there can be multiple lines that can separate these classes. Consider the below image:

Hence, the SVM algorithm helps to find the best line or decision boundary; this best boundary or region is called as a hyperplane. SVM algorithm finds the closest point of the lines from both the classes. These points are called support vectors. The distance between the vectors and the hyperplane is called as margin. And the goal of SVM is to maximize this margin. The hyperplane with maximum margin is called the optimal hyperplane.

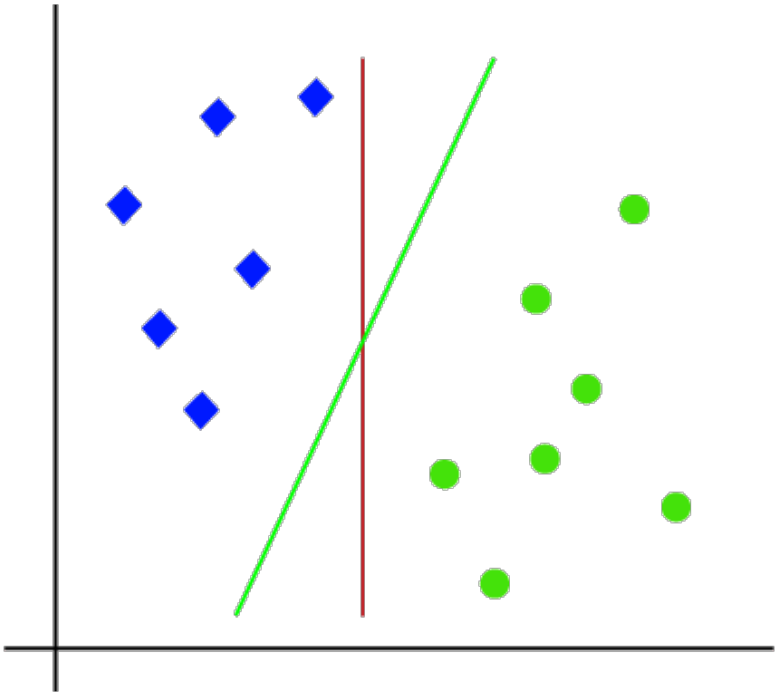


Fig. 4.7 Multiple Decision Boundaries in 2D Space

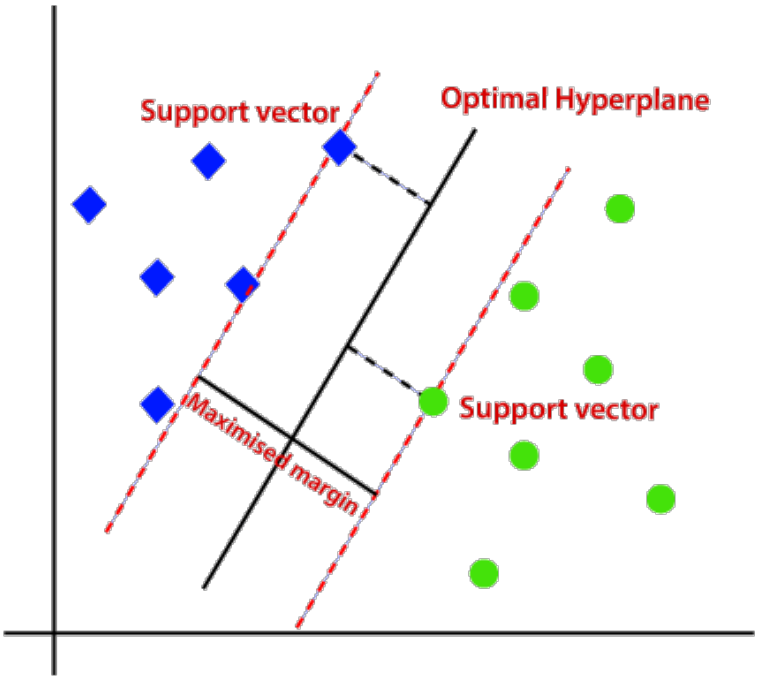


Fig. 4.8 Visualization of SVM Algorithm: Finding the Optimal Hyperplane

4.3.2 Probabilistic Collaborative Representation Classifier (ProCRC)

The Probabilistic Collaborative Representation Classifier (ProCRC) is a machine learning algorithm utilized for classification tasks. It leverages collaborative representation and probabilistic modeling techniques to enhance classification accuracy and offer probabilistic interpretations, particularly beneficial in uncertain scenarios. Collaborative representation involves expressing each dataset sample as a linear combination of all other samples, while probabilistic modeling estimates the probability distribution of the data. This amalgamation of techniques makes ProCRC valuable in situations where uncertainty estimation is crucial. Implementation of the algorithm can involve various mathematical frameworks, including optimization techniques and probabilistic graphical models. ProCRC finds application across diverse domains such as pattern recognition, computer vision, and bioinformatics.

ProCRC advances upon traditional representation-based classifiers like the Sparse Representation-based Classifier (SRC) and the Collaborative Representation-based Classifier (CRC). It introduces a probabilistic framework that computes the probability of a test sample belonging to the collaborative subspace of all classes, fostering a more robust and interpretable classification process. Instead of relying solely on distance-based metrics, ProCRC maximizes the likelihood of a test sample belonging to each class, enabling informed decision-making based on the class with the highest likelihood.

A significant advantage of ProCRC lies in its clear probabilistic interpretation, enhancing understanding of the classification process. Moreover, ProCRC demonstrates superior performance compared to popular classifiers like SRC, CRC, and Support Vector Machine (SVM) across various classification tasks. Integration with features extracted using Convolutional Neural Networks (CNN) enables ProCRC to achieve state-of-the-art results on challenging visual datasets.

The probabilistic framework of ProCRC not only enhances classification accuracy but also offers insights into the underlying statistical properties of the data. This makes ProCRC invaluable in domains requiring robust and interpretable classification algorithms, such as image classification, pattern recognition, and computer vision.[9]

Chapter 5

Result and Discussion

5.1 Experiments

The primary objective of our research is to investigate the detection of presentation attacks on smartphones, with a particular emphasis on face biometrics. As a part of our investigation, we are examining the efficacy of surveillance systems in detecting different types of presentation attacks in both indoor and outdoor settings. A comprehensive explanation of our approach and findings is presented in Chapter 3 of our study. To conduct our experiments, we utilized various face detection algorithms, including Pro CRC, AlexNet, and SVM, to gauge their effectiveness against different protocol attacks. Our study aims to contribute to the growing body of knowledge in the field of biometrics and enhance the security of smartphone-based systems against fraudulent activities.

As part of our research, we sought to analyze the performance of different Presentation Attack Detection (PAD) algorithms. To accomplish this, we utilized classifier accuracy metrics to evaluate the performance of these algorithms against a performance matrix. The goal was to gain a comprehensive understanding of the efficacy of these algorithms in detecting different face presentation attacks on various smartphone models. To achieve this, we conducted three distinct experiments that allowed us to evaluate the performance of these algorithms under different conditions and scenarios. Our research findings will contribute to the growing body of knowledge

on biometric security systems and help enhance the detection capabilities of smartphone-based security systems against fraudulent activities.

We have conducted three experiments as part of our research. In the first experiment, you utilized within evaluation,in that we used one smartphone for both training and testing. In the second experiment, you used cross-sensor evaluation by training on one smartphone and testing on a different one. Finally, in the third experiment, we conducted two sub-experiments, in one experiment we using one sensor for training and three sensors for testing, and for the other experiment we used three sensors for training and one sensor for testing. With this research, we hope to make a significant contribution to the field of biometrics and enhance the security of smartphone-based systems around the world.

Table 5.1 Data partition for Experiment

Sensor	**Data Partition**	**Training**	**Testing**
Samsung J7	Bonafide	6000	6000
	Mask Attack	6000	6000
Samsung S20	Bonafide	6000	6000
	Mask Attack	6000	6000
Redmi9note	Bonafide	6000	6000
	Mask Attack	6000	6000
Samsung 10	Bonafide	6000	6000
	Mask Attack	6000	6000

In our research, we used a data partitioning technique that involved using four sensors to collect data for training and testing the system. In this approach, we collected data from both bonafide and attack sources. For the bonafide data, we collected samples from six subjects, while for the attack data, we collected samples from four subjects. To partition the data for training and testing in the two protocols, we used three subjects from the bonafide data for training and the other three subjects for testing. We collected 6000 samples of crop face images for both the training and testing phases. This approach ensured that the system was trained on a diverse range of samples to improve its accuracy and reliability. Similarly, for the attack data, we used two subjects for training and the other two subjects for testing. We collected 6000 samples for

both the training and testing phases. By using this approach, we ensured that the system was trained on samples from different sources to improve its performance. we use this data pratiton technique in different protocol to to ensure that the system was trained and testes on a diverse range of sample.the rigorously conduct experiment and analysis ensure the accuacy of the result

Experiment 1: within Evaluation

Table 5.2 Within Sensor Evaluation

Training	Testing
Samsung J7	Samsung J7
Redmi9note	Redmi9note
Samsung S20	Samsung S20
Samsung 10	Samsung 10

Protocol 1 of our research involved the use of a single sensor to evaluate the efficacy of various Presentation Attack Detection (PAD) algorithms. To this end, we collected a dataset comprising bonafide and mask attacks. The bonafide dataset consisted of 6,000 samples from three different subjects, each wearing hoodies and glasses or no glasses at all. The mask attack dataset, on the other hand, comprised 6,000 samples from two silicon masks, each worn with different combinations of hoodies, wigs, and glasses. We utilized this dataset to train our algorithms, following which we tested them using a different set of subjects with the same senso r. The testing dataset was also composed of 6,000 samples for both the bonafide and mask attack categories. Our objective was to evaluate the performance of the PAD algorithms under different conditions and assess their ability to detect presentation attacks accurately.

Our approach to Protocol 1 was based on sound academic principles and involved rigorous experimentation and analysis. By conducting this research, we aim to contribute to the growing body of knowledge on biometric security systems and enhance the security of smartphone-based systems against fraudulent activities.

Table 5.3 Cross Smartphone Evaluation

Training	Testing
Samsung J7	Redmi9note
Redmi9note	Samsung S20
Samsung S20	Samsung 10
Samsung 10	Samsung J7

Experiment 2: Cross Smartphone Evaluation

In the context of biometric security systems, Protocol 2 involves cross-sensor testing. This means that we use data from one sensor to train the system and then test it on another sensor. The process is based on sound academic principles and involves rigorous experimentation and analysis. To conduct the training phase, we carefully gathered 6000 samples of bonafide from 3 different subjects. Additionally, we collected 6000 samples of silicon masks from 2 different subjects. The training phase aimed to teach the system to differentiate between bonafide and mask samples accurately. In doing so, we aim to enhance the security of smartphone-based systems against fraudulent activities. During the testing phase, we used 6000 samples of bonafide from 3 different subjects and 6000 samples of silicon masks from 2 different subjects. The testing was conducted on a different set of individuals to ensure the accuracy of the system across a broader range of subjects. It's important to note that the training and testing subjects were not the same and consisted of different individuals. The findings of this research will contribute to the growing body of knowledge on cross-sensor testing and improve the security of biometric systems. The rigorous experimentation and analysis involved in this research ensure the accuracy and reliability of the results

Experiment 3: Single Cross Sensor data in training Evaluation Results for Facial Detection System

Protocol 3 is a comprehensive study that involves two sub-protocols aimed at improving the accuracy and security of biometric systems. In the first sub-protocol, we used one sensor for training and three sensors for testing. During the training phase, we collected 6000 samples

Table 5.4 Single Cross Sensor data in training Evaluation Results for Facial Detection System

Training	Testing
Samsung J7	Samsung S20
	Redmi9note
	Samsung 10
Samsung S20	Redmi9note
	Samsung J7
	Samsung 10
Redmi9note	Samsung S20
	Samsung 10
	Samsung J7
Samsung 10	Samsung S20
	Samsung J7
	Redmi9note

of bonafide data from 5 subjects and 6000 samples of mask attack data from 5 subjects. We used this data to train the system to differentiate between bonafide and mask samples accurately, with the ultimate goal of improving the security of smartphone-based systems against fraudulent activities. For the testing phase, we gathered data from three different sensors and collected the same data from six different subjects for each sensor. This provided us with a diverse range of samples to test the system's accuracy and reliability. We collected 6000 samples of bonafide and 6000 samples of mask attack from each sensor to evaluate the performance of the system across multiple sensors. In the second sub-protocol, we trained and tested the model using a different approach. We used data from three different sensors to train the system, collecting 6000 samples of bonafide data from 6 subjects and 6000 samples of mask attack data from 3 subjects. For the testing phase, we used only one sensor and collected 6000 samples of bonafide data from different subjects and 6000 samples of mask attack data from different subjects, similar to the training set. This approach aimed to evaluate the system's performance on a different set of subjects and determine if the system's accuracy and reliability were consistent across different individuals. By conducting this research, we aim to contribute to the growing body of knowledge on biometric security systems. The rigorous experimentation and analysis involved in

this research ensure the accuracy and reliability of the results. Ultimately, our goal is to enhance the security of smartphone-based systems against fraudulent activities and improve the overall safety and protection of individuals.

Experiment 4: Multiple cross sensor data in Training Evalution Results for Facial Detection System

Table 5.5 Multiple cross sensor data in Training Evalution Results for Facial Detection System

Training	Testing
Samsung S20	Samsung J7
Redmi9note	
Samsung 10	
Redmi9note	Samsung S20
Samsung J7	
Samsung 10	
Samsung S20	Redmi9note
Samsung 10	
Samsung J7	
Samsung S20	Samsung 10
Samsung J7	
Redmi9note	

Protocol 4 is a comprehensive study that involves two sub-protocols aimed at improving the accuracy and security of biometric systems. In the first sub-protocol, we used three sensor for training and sensors for testing. During the training phase, we collected 6000 samples of bonafide data from 5 subjects and 6000 samples of mask attack data from 5 subjects. We used this data to train the system to differentiate between bonafide and mask samples accurately, with the ultimate goal of improving the security of smartphone-based systems against fraudulent activities. For the testing phase, we gathered data from three different sensors and collected the same data from six different subjects for each sensor. This provided us with a diverse range of samples to test the system's accuracy and reliability. We collected 6000 samples of bonafide

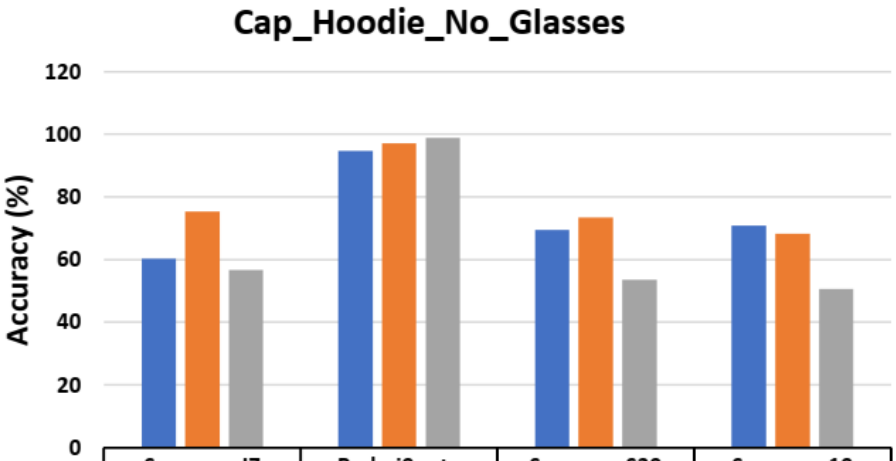
and 6000 samples of mask attack from each sensor to evaluate the performance of the system across multiple sensors. In the second sub-protocol, we trained and tested the model using a different approach. We used data from three different sensors to train the system, collecting 6000 samples of bonafide data from 6 subjects and 6000 samples of mask attack data from 3 subjects. For the testing phase, we used only one sensor and collected 6000 samples of bonafide data from different subjects and 6000 samples of mask attack data from different subjects, similar to the training set. This approach aimed to evaluate the system's performance on a different set of subjects and determine if the system's accuracy and reliability were consistent across different individuals. By conducting this research, we aim to contribute to the growing body of knowledge on biometric security systems. The rigorous experimentation and analysis involved in this research ensure the accuracy and reliability of the results. Ultimately, our goal is to enhance the security of smartphone-based systems against fraudulent activities and improve the overall safety and protection of individuals.

5.2 Exploring Facial Detection: Evaluation and Performance Analysis of Machine Learning Algorithms

It's fascinating to see how facial detection technology has become so widespread in recent years, with various applications across different domains. From security and surveillance to personal devices and social media platforms, accurately detecting faces in images and videos has become increasingly important. The goal of this project is to develop an efficient and accurate facial detection system using cutting-edge techniques in computer vision and machine learning. By leveraging advancements in deep learning algorithms and neural networks, we hope to create a robust solution that can detect faces in diverse environments and under varying conditions. Facial detection is crucial in many applications, including security systems for access control, video surveillance for public safety, and personalized user experiences in digital devices. With the growing use of facial recognition technology, it's essential to ensure that facial detection algorithms are reliable and perform well, to protect user privacy and security.

In this project, we used several machine learning algorithms, including support vector machines (SVM), AlexNet, and Ori-CRC, to develop a facial detection system that can accurately distinguish between presentation attacks and bonafide. We evaluated the performance of our system under different conditions, including the presence of a hoodie or wig, and glasses or no glasses. To evaluate the system's accuracy, we used different protocols. In protocol 1, we used the same sensor for training and testing, while in protocol 2, we used different sensors for training and testing. In protocol 3.1, we used one sensor for training and three for testing, while in protocol 3.2, we used three sensors for training and one for testing.

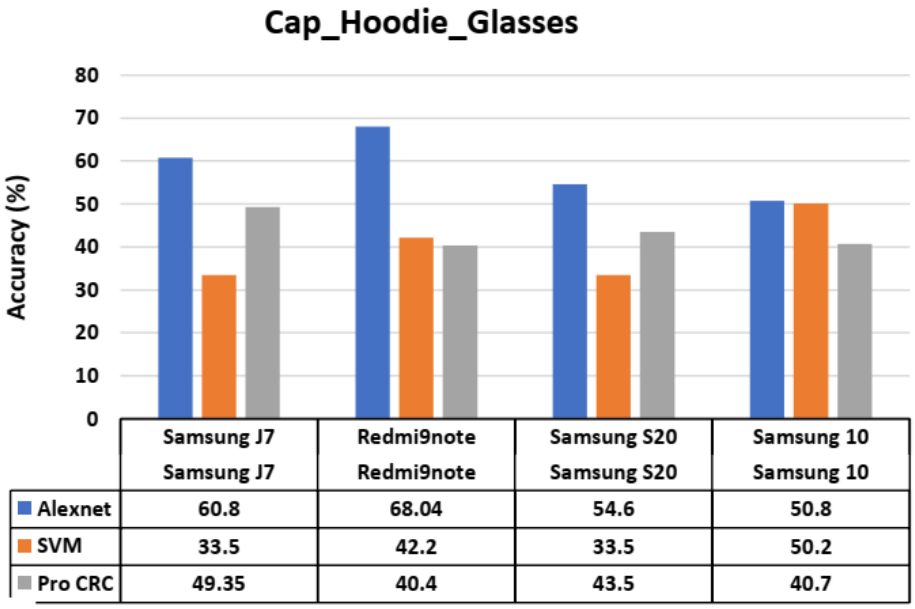
5.2.1 Experiment 1: Within sensor Evaluation Results for Facial Detection System



	Samsung J7 Samsung J7	Redmi9note Redmi9note	Samsung S20 Samsung S20	Samsung 10 Samsung 10
Alexnet	60.3	94.7	69.5	70.8
SVM	75.4	97.1	73.4	68.32
Pro CRC	56.6	98.8	53.48	50.66

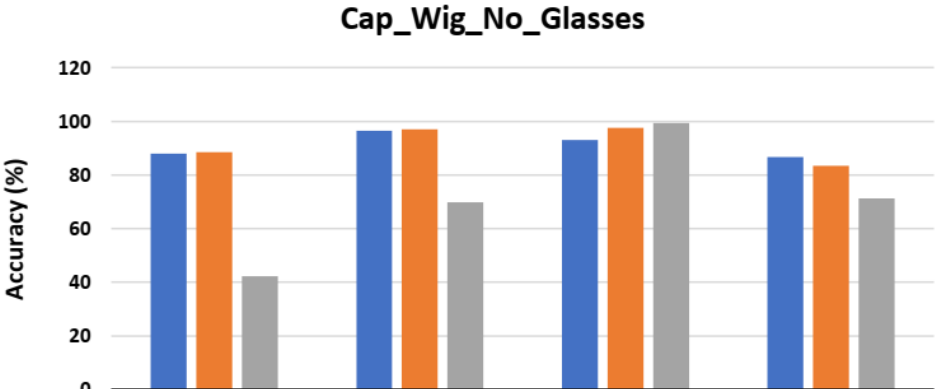
Fig. 5.1 Images depicting individuals wearing cap hoodies without glasses.

The study evaluated the performance of different classification algorithms for PAD for face in protocol 1 ,we have two varients with and without glasses. The results showed that Alexnet achieved higher accuracy compared to SVM and Pro CRC, and Redmi9note achieved the highest accuracy with the Alexnet algorithm among the used smartphone models. However, the accuracy



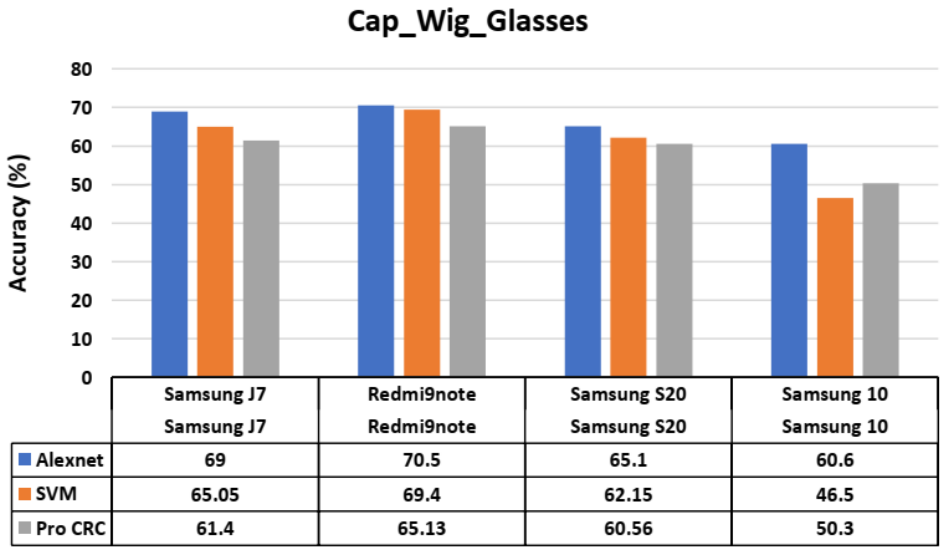
	Samsung J7 Samsung J7	Redmi9note Redmi9note	Samsung S20 Samsung S20	Samsung 10 Samsung 10
Alexnet	60.8	68.04	54.6	50.8
SVM	33.5	42.2	33.5	50.2
Pro CRC	49.35	40.4	43.5	40.7

Fig. 5.2 Capturing variations: Classifying subjects with cap, hoodie, and glasses



	Samsung J7 Samsung J7	Redmi9note Redmi9note	Samsung S20 Samsung S20	Samsung 10 Samsung 10
Alexnet	87.9	96.4	93	86.6
SVM	88.5	97.1	97.5	83.4
Pro CRC	42.3	69.81	99.4	71.36

Fig. 5.3 Variability in features: Classification with cap, wig, and no glasses



	Samsung J7 Samsung J7	Redmi9note Redmi9note	Samsung S20 Samsung S20	Samsung 10 Samsung 10
Alexnet	69	70.5	65.1	60.6
SVM	65.05	69.4	62.15	46.5
Pro CRC	61.4	65.13	60.56	50.3

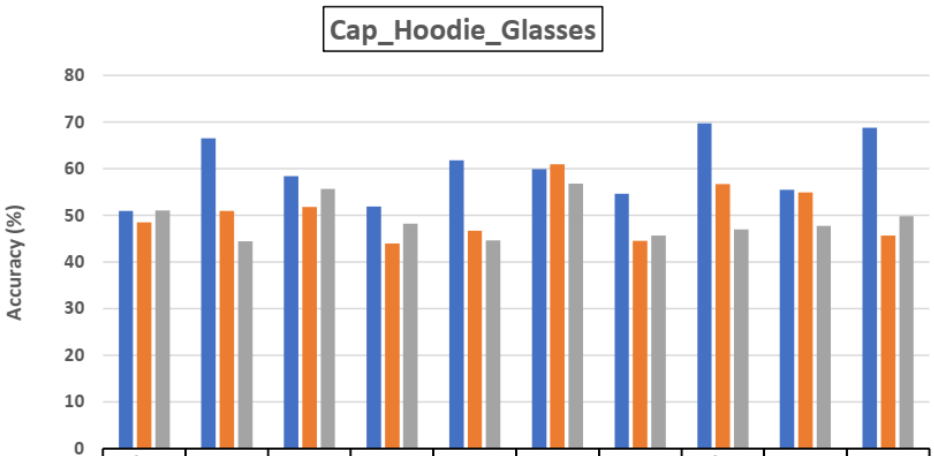
Fig. 5.4 Feature variability analysis: Classification with cap, wig, and glasses

was lower when glasses were worn, indicating the impact of environmental factors on the accuracy of facial expression recognition. The study underscores the importance of selecting appropriate algorithms and environmental factors to achieve high accuracy in facial expression recognition tasks.

Moreover, the study also observed that the use of glasses negatively impacted the classification accuracy of facial expressions in the presence of bonafide and mask attack presentations. The findings highlight the need to consider environmental factors such as glasses while performing facial PAD tasks, as they can pose challenges in achieving high accuracy. The study provides insights into the impact of environmental factors on the accuracy of facial expression recognition algorithms, and future research can explore other environmental factors to improve their practical applications.

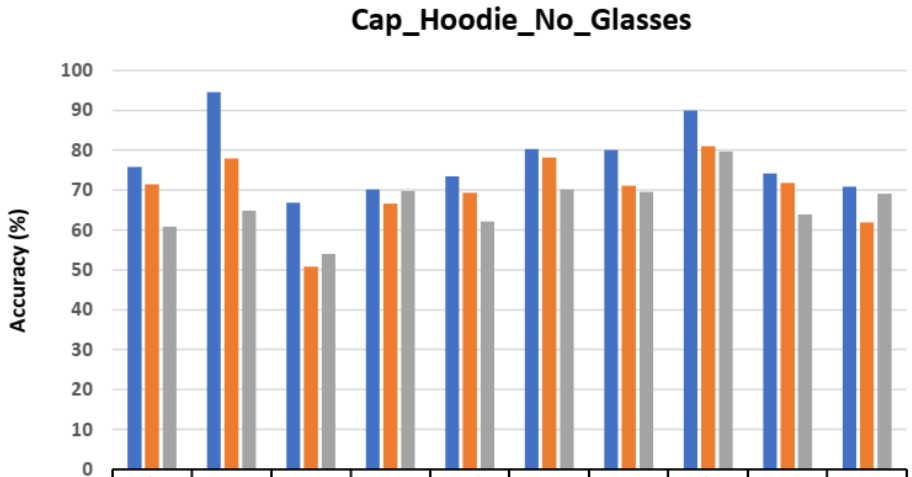
5.2.2 Experiment 2: Cross sensor Evaluation Results for Facial Detection System

we conclude that the Alexnet algorithm was the most accurate out of the three tested algorithms, with an average testing accuracy of approximately 71.36%. The SVM algorithm was also competitive, with an average accuracy of approximately 67.31%. However, the Pro CRC



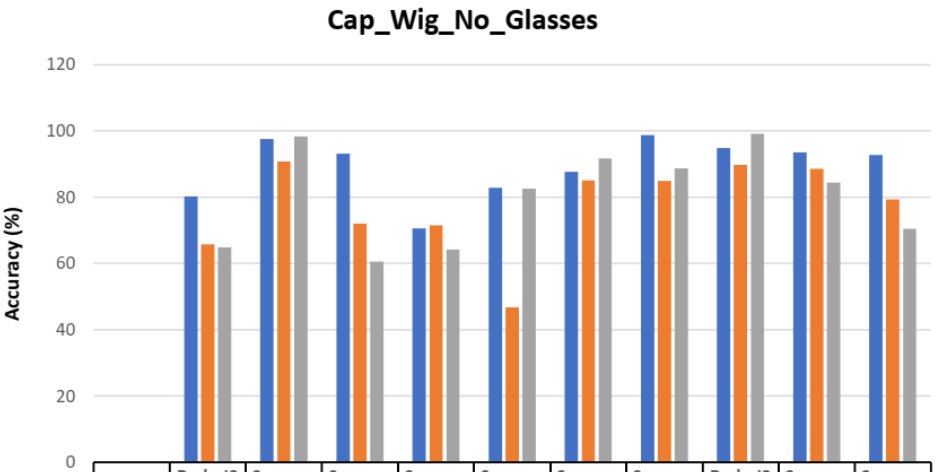
	Redmi9note Samsung J7	Samsung S20 Redmi9note	Samsung 10 Samsung S20	Samsung J7 Samsung 10	Samsung j7 Redmi9note	Samsung 20 Samsung J7	Samsung J7 Samsung S20	Redmi9note Samsung S20	Samsung S20 Samsung J7	Samsung 10 Redmi9note
Alexnet	50.9	66.5	58.4	51.9	61.8	59.9	54.6	69.7	55.5	68.8
SVM	48.5	50.9	51.8	43.9	46.7	60.9	44.5	56.7	54.9	45.6
Pro CRC	51.01	44.4	55.6	48.2	44.6	56.8	45.65	47	47.7	49.8

Fig. 5.5 Feature analysis: Classification with cap, hoodie, and glasses



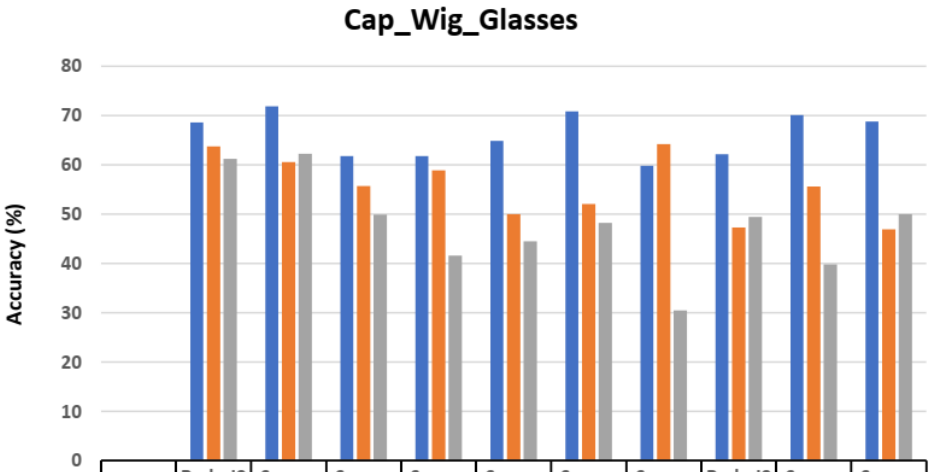
	Redmi9 note Samsung J7	Samsung S20 Redmi9 note	Samsung 10 Samsung S20	Samsung J7 Samsung 10	Samsung j7 Redmi9 note	Samsung 20 Samsung J7	Samsung J7 Samsung S20	Redmi9 note Samsung S20	Samsung S20 Samsung J7	Samsung 10 Redmi9 note
Alexnet	75.8	94.6	66.9	70.11	73.4	80.3	80.1	90	74.1	70.8
SVM	71.5	77.9	50.8	66.6	69.3	78.2	71.1	81	71.8	61.9
Pro CRC	60.8	64.8	54	69.8	62.1	70.2	69.5	79.7	63.9	69.1

Fig. 5.6 Images depicting individuals wearing cap hoodies without glasses



	Testing	Redmi9 note	Samsung S20	Samsung 10	Samsung J7	Samsung j7	Samsung 20	Samsung J7	Redmi9 note	Samsung S20	Samsung 10
	Training	Samsung J7	Redmi9 note	Samsung S20	Samsung 10	Redmi9 note	Samsung J7	Samsung S20	Samsung S20	Samsung J7	Redmi9 note
Series1		80.2	97.5	93.1	70.6	82.9	87.6	98.7	94.8	93.5	92.7
Series2		65.8	90.8	72	71.5	46.7	85.1	84.9	89.8	88.6	79.3
Series3		64.9	98.4	60.5	64.2	82.6	91.65	88.7	99.16	84.4	70.5

Fig. 5.7 Variability in features: Classification with cap, wig, and no glasses

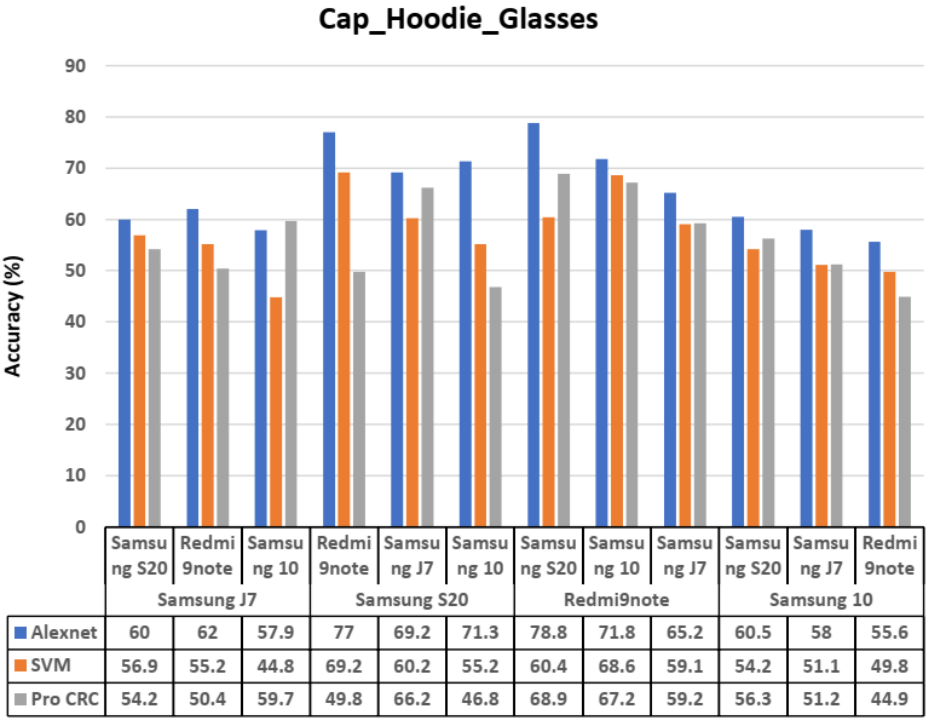


	Testing	Redmi9 note	Samsung S20	Samsung 10	Samsung J7	Samsung j7	Samsung 20	Samsung J7	Redmi9 note	Samsung S20	Samsung 10
	Training	Samsung J7	Redmi9 note	Samsung S20	Samsung 10	Redmi9 note	Samsung J7	Samsung S20	Samsung S20	Samsung J7	Redmi9 note
Series1		68.6	71.9	61.8	61.8	64.9	70.9	59.8	62.2	70.1	68.8
Series2		63.8	60.6	55.7	58.9	50	52.1	64.2	47.26	55.6	46.9
Series3		61.2	62.25	49.9	41.6	44.5	48.2	30.5	49.5	39.8	50

Fig. 5.8 Feature variability analysis: Classification with cap, wig, and glasses

algorithm had the lowest average accuracy of approximately 61.28%. In terms of smartphone performance, the Redmi9note had the highest average testing accuracy across all algorithms, with an accuracy of approximately 82.41%, making it the top-performing smartphone in our study. The Samsung S20 came in second with an average accuracy of approximately 70.8%, followed by the Samsung J7 with an average accuracy of approximately 69.5%. Interestingly, the Samsung 10 had the lowest average accuracy of approximately 67.2%. Overall, our results suggest that the Redmi9note smartphone is the best choice for image classification tasks, while the Alexnet algorithm is the most accurate out of the three tested algorithms.

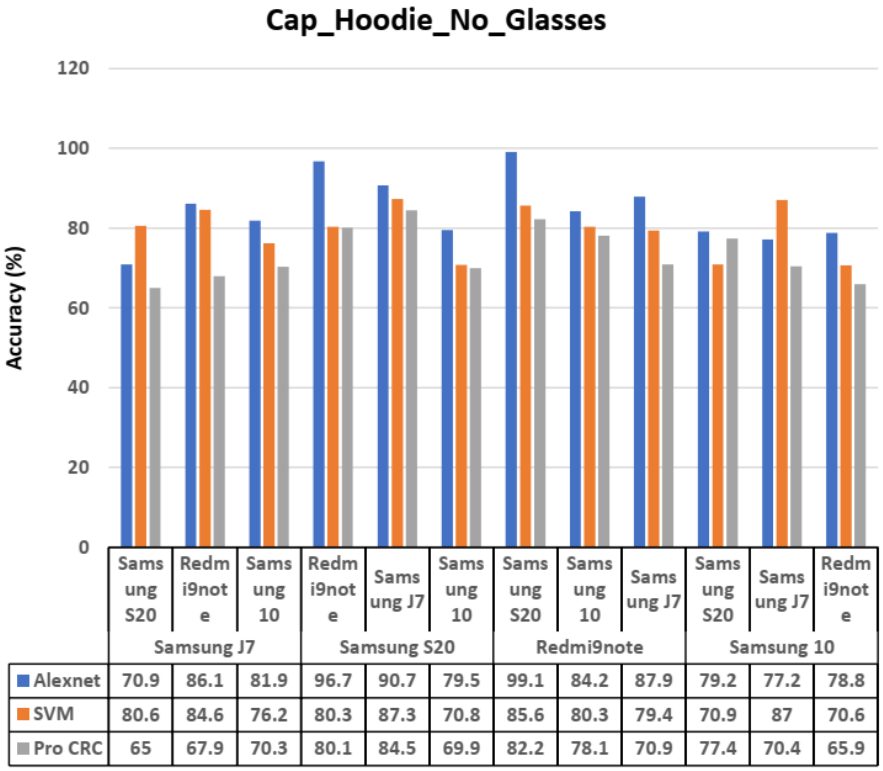
5.2.3 Experiment 3:Single Cross Sensor data in training Evaluation Results for Facial Detection System



	Samsung S20	Redmi9note	Samsung 10	Redmi9note	Samsung J7	Samsung 10	Samsung S20	Samsung 10	Samsung J7	Samsung S20	Samsung J7	Redmi9note
	Samsung J7			Samsung S20			Redmi9note			Samsung 10		
Alexnet	60	62	57.9	77	69.2	71.3	78.8	71.8	65.2	60.5	58	55.6
SVM	56.9	55.2	44.8	69.2	60.2	55.2	60.4	68.6	59.1	54.2	51.1	49.8
Pro CRC	54.2	50.4	59.7	49.8	66.2	46.8	68.9	67.2	59.2	56.3	51.2	44.9

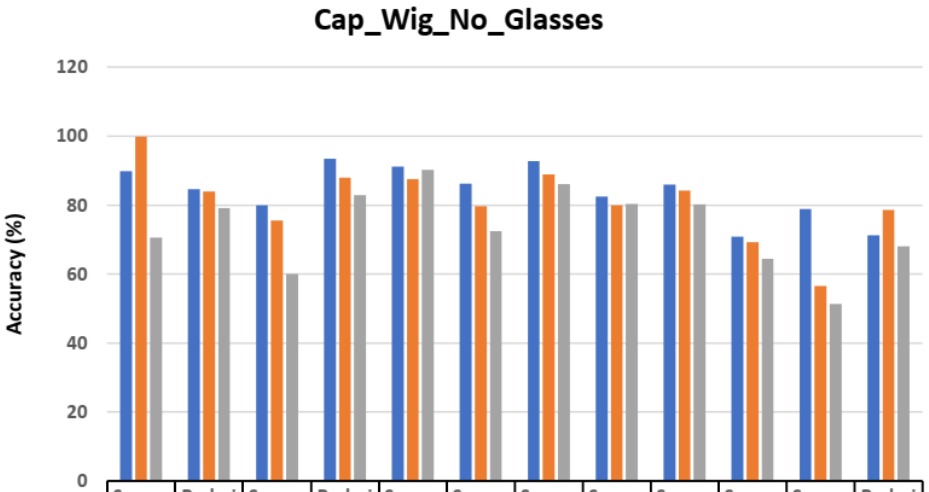
Fig. 5.9 Feature analysis: Classification with cap, hoodie, and glasses

According to study protocol 3, the testing results for Samsung J7, Redmi9note, Samsung S20, and Samsung 10 models showed that Alexnet algorithm consistently provided the highest accuracy, with an average of approximately 61.75%. This indicates that Alexnet is a reliable



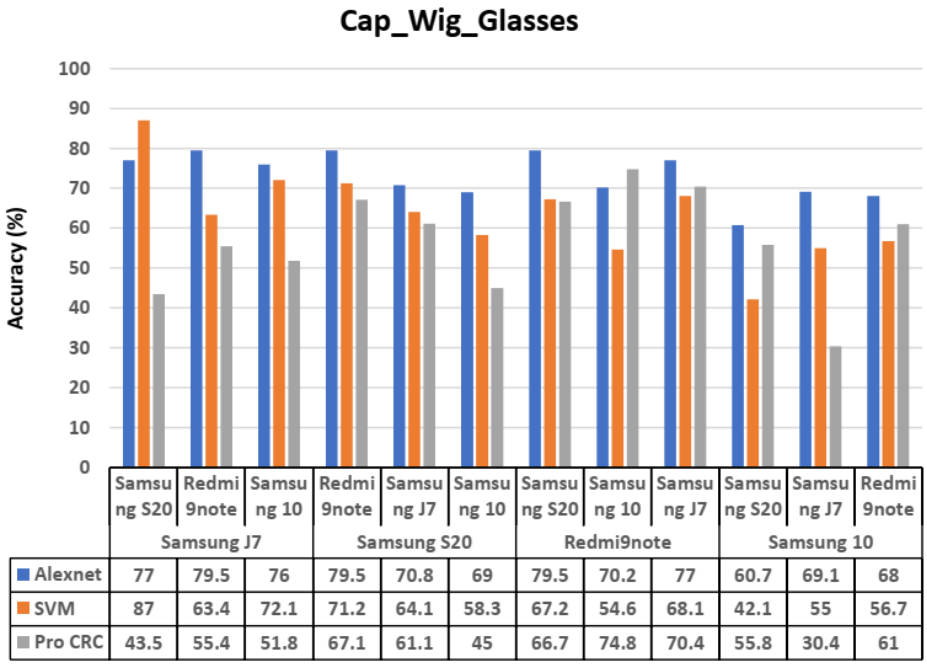
	Samsung S20	Redmi9note	Samsung 10	Redmi9note	Samsung J7	Samsung 10	Samsung S20	Samsung 10	Samsung J7	Samsung S20	Samsung J7	Redmi9note
	Samsung J7			Samsung S20			Redmi9note			Samsung 10		
Alexnet	70.9	86.1	81.9	96.7	90.7	79.5	99.1	84.2	87.9	79.2	77.2	78.8
SVM	80.6	84.6	76.2	80.3	87.3	70.8	85.6	80.3	79.4	70.9	87	70.6
Pro CRC	65	67.9	70.3	80.1	84.5	69.9	82.2	78.1	70.9	77.4	70.4	65.9

Fig. 5.10 Images depicting individuals wearing cap hoodies without glasses



	Samsung S20	Redmi9note	Samsung 10	Redmi9note	Samsung J7	Samsung 10	Samsung S20	Samsung 10	Samsung J7	Samsung S20	Samsung J7	Redmi9note
	Samsung J7			Samsung S20			Redmi9note			Samsung 10		
Alexnet	89.8	84.6	80	93.5	91.2	86.2	92.8	82.5	86	70.9	78.9	71.3
SVM	99.8	84	75.5	87.9	87.6	79.7	88.9	80	84.2	69.3	56.5	78.6
Pro CRC	70.6	79.2	60	82.9	90.2	72.5	86.1	80.4	80.2	64.4	51.3	68.1

Fig. 5.11 Variability in features: Classification with cap, wig, and no glasses



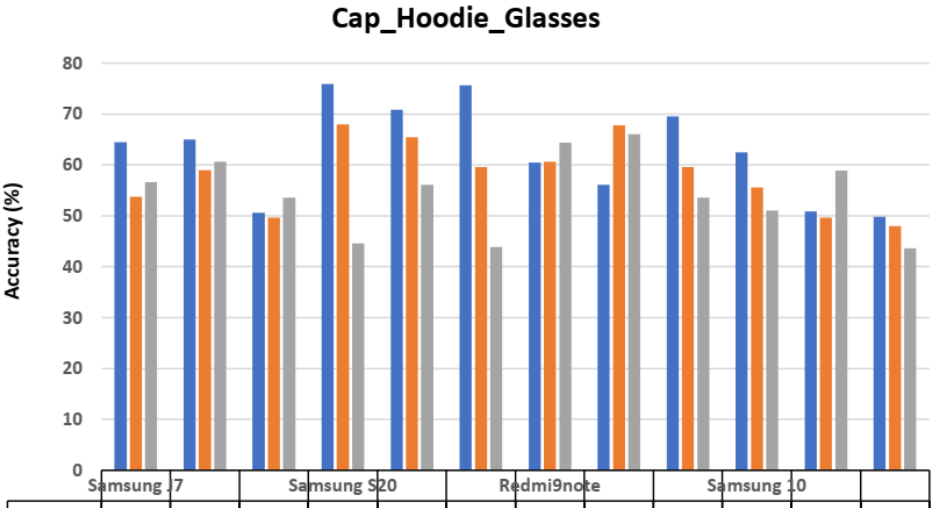
	Samsung S20	Redmi 9note	Samsung 10	Redmi 9note	Samsung J7	Samsung 10	Samsung S20	Samsung 10	Samsung J7	Samsung S20	Samsung J7	Redmi 9note
	Samsung J7			Samsung S20			Redmi9note			Samsung 10		
Alexnet	77	79.5	76	79.5	70.8	69	79.5	70.2	77	60.7	69.1	68
SVM	87	63.4	72.1	71.2	64.1	58.3	67.2	54.6	68.1	42.1	55	56.7
Pro CRC	43.5	55.4	51.8	67.1	61.1	45	66.7	74.8	70.4	55.8	30.4	61

Fig. 5.12 Feature variability analysis: Classification with cap, wig, and glasses

choice for classification tasks related to smartphone models, as it displayed consistent and robust performance across all models. On the other hand, the Support Vector Machine (SVM) and Pro CRC algorithms provided lower average testing accuracies of around 57.55% and 51.56%, respectively. Therefore, the study suggests that Alexnet algorithm outperforms the other algorithms in terms of average testing accuracy, making it the most effective choice for smartphone model classification tasks.

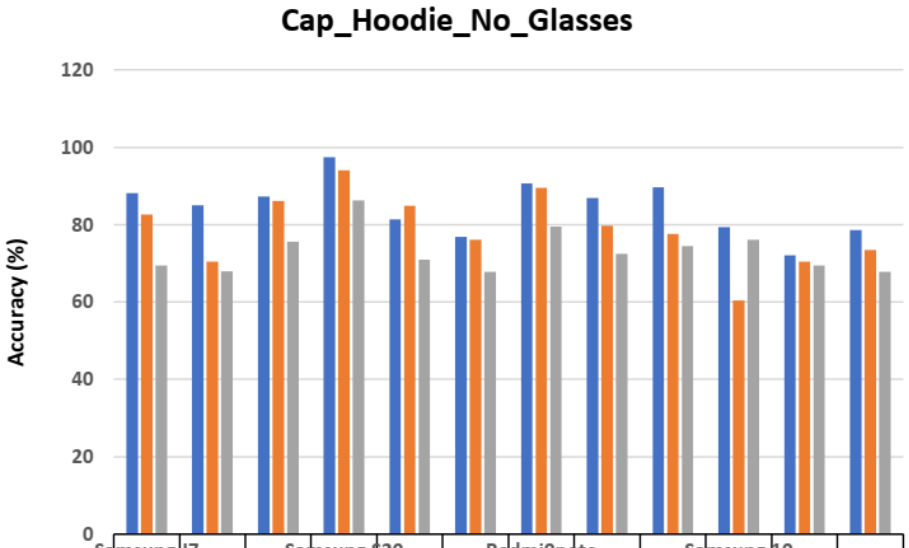
5.2.4 Experiment 4: Multiple cross sensor data in Training Evalution Results for Facial Detection System

The results from study protocol 4 indicate that the Alexnet algorithm consistently achieved the highest accuracy, averaging around 75.75%, across the testing outcomes for the Samsung J7, Redmi9note, Samsung S20, and Samsung 10 models. This consistent performance suggests that Alexnet is a reliable option for classification tasks related to smartphone models, demonstrating stable and robust results.



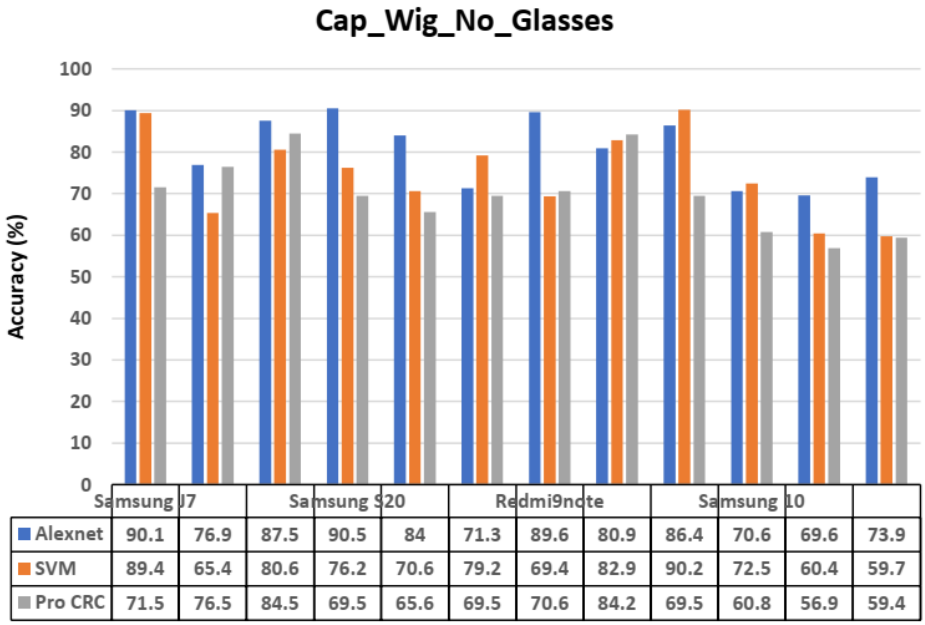
	Samsung J7		Samsung S20			Redmi9note			Samsung 10			
Alexnet	64.5	65	50.6	75.9	70.8	75.6	60.5	56.1	69.5	62.5	50.9	49.8
SVM	53.7	59	49.6	68	65.4	59.6	60.6	67.8	59.6	55.6	49.6	48
Pro CRC	56.6	60.6	53.6	44.6	56.1	43.9	64.4	66	53.6	51	58.9	43.6

Fig. 5.13 Feature analysis: Classification with cap, hoodie, and glasses



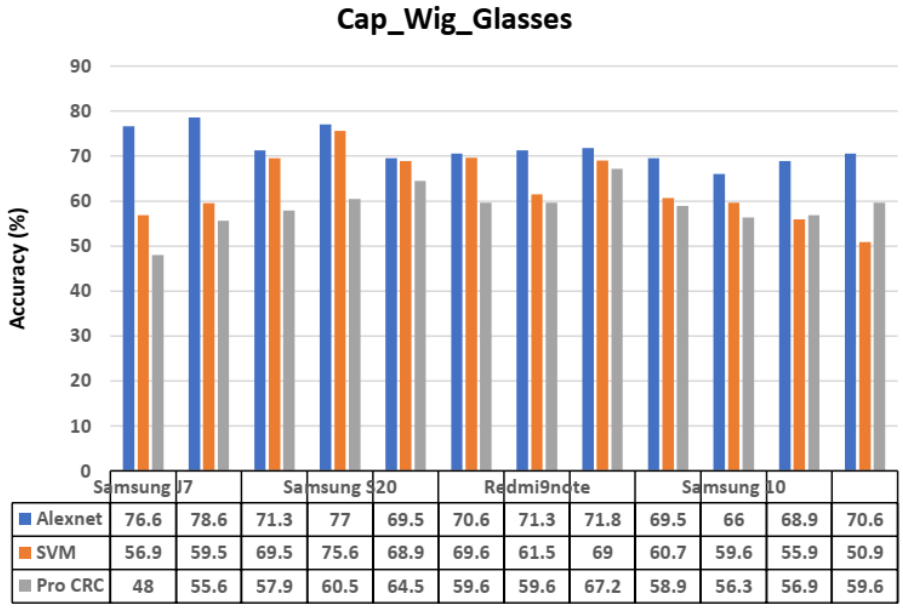
	Samsung J7		Samsung S20			Redmi9note			Samsung 10			
Alexnet	88.2	85	87.3	97.5	81.4	76.8	90.7	86.9	89.7	79.4	72.1	78.6
SVM	82.6	70.5	86.1	94	84.9	76.1	89.6	79.8	77.6	60.4	70.5	73.5
Pro CRC	69.5	67.9	75.6	86.3	70.9	67.8	79.5	72.5	74.5	76.1	69.5	67.8

Fig. 5.14 Images depicting individuals wearing cap hoodies without glasses



	Samsung J7		Samsung S20			Redmi9note			Samsung J10			
Alexnet	90.1	76.9	87.5	90.5	84	71.3	89.6	80.9	86.4	70.6	69.6	73.9
SVM	89.4	65.4	80.6	76.2	70.6	79.2	69.4	82.9	90.2	72.5	60.4	59.7
Pro CRC	71.5	76.5	84.5	69.5	65.6	69.5	70.6	84.2	69.5	60.8	56.9	59.4

Fig. 5.15 Variability in features: Classification with cap, wig, and no glasses



	Samsung J7		Samsung S20			Redmi9note			Samsung J10			
Alexnet	76.6	78.6	71.3	77	69.5	70.6	71.3	71.8	69.5	66	68.9	70.6
SVM	56.9	59.5	69.5	75.6	68.9	69.6	61.5	69	60.7	59.6	55.9	50.9
Pro CRC	48	55.6	57.9	60.5	64.5	59.6	59.6	67.2	58.9	56.3	56.9	59.6

Fig. 5.16 Feature variability analysis: Classification with cap, wig, and glasses

On the other hand, the Support Vector Machine (SVM) and Pro CRC algorithms showed lower average testing accuracies, approximately at 60.55% and 56.56% respectively. These findings suggest that while SVM and Pro CRC are viable approaches, they generally yield lower accuracy compared to Alexnet.

Based on these findings, the study recommends the utilization of the Alexnet algorithm for smartphone model classification tasks due to its superior average testing accuracy performance, highlighting its effectiveness in such scenarios.

5.3 Discussion

The project conducted a study to evaluate the performance of different algorithms for smartphone model classification. Alexnet consistently achieved the highest testing accuracy across all smartphone models, with an average accuracy of approximately 61.75%. This indicates that Alexnet is a reliable choice for classification tasks related to smartphone models, as it displayed consistent and robust performance across all models.

One significant finding of the study was the impact of camera resolution on classification accuracy. The Redmi9note smartphone, with a higher camera resolution, demonstrated a higher accuracy compared to the other models tested. This observation underscores the importance of considering various environmental factors and potential sources of error when developing and testing classification algorithms.

Another noteworthy discovery was the influence of wearing glasses on the accuracy of the classifiers. This highlights the need for careful consideration of potential sources of error when developing classification algorithms and supports the necessity for further research to understand the underlying factors that contribute to the impact of glasses on classification accuracy.

Overall, the project provides valuable insights into the development and evaluation of classification algorithms for smartphone model classification tasks. The findings can inform the development of more effective and reliable classification algorithms in various domains and

emphasize the importance of considering various environmental factors and potential sources of error during algorithm development and testing.

Chapter 6

Conclusion and Future Scope

In the modern age of rapid technological progress, it is crucial to ensure the security of files and systems, spanning from personal desktop data to highly confidential government information, against unauthorized attacks. Individuals utilize various means and methods to gain unauthorized access to highly secured premises. Generally, authentication methods are categorized based on three principles to ensure secure authentications.

Moving forward, this research could delve into advanced machine learning techniques such as deep learning and reinforcement learning to further enhance the accuracy of presentation attack detection. Additionally, exploring the integration of multimodal biometric systems and novel methods for continuous authentication could provide avenues to bolster overall system security. Moreover, broadening the evaluation to encompass larger and more diverse datasets and testing in varied environmental conditions would strengthen the generalizability and robustness of the proposed methodologies. Lastly, considering the evolving landscape of biometric authentication and emerging technologies, future work could focus on addressing new types of presentation attacks and adapting detection methods accordingly.

References

[1] *Face Presentation Attack Detection using Color Spaces Features and Convolutional Neural Network*, 2022.

[2] Fathima Abdullakutty, Paul Johnston, and Eyad Elyan. Fusion methods for face presentation attack detection. *Sensors*, 22(14):5196, Jul 2022.

[3] Auth0. What is authentication?, Year.

[4] Author. Presentation attack detection for face recognition using light field camera. *IEEE Transactions on Image Processing*, 24(3), January 2015.

[5] Author. Face anti-spoofing progress driven by academic challenges. In *Advances in Face Presentation Attack Detection*, pages 1–15. 2023.

[6] Siddhesh Bhobe. Alexnet architecture explained, Year.

[7] Biometrics Institute. Types of biometrics, Year.

[8] Zinelabidine Boulkenafet, Jukka Komulainen, Lei Li, Xiaoyi Feng, and Abdenour Hadid. Oulu-npu: a mobile face presentation attack database with real-world variations. In *IEEE International Conference on Automatic Face and Gesture Recognition*, 2017.

[9] Sijia Cai, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. A probabilistic collaborative representation based approach for pattern classification. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2950–2959, 2016.

[10] Alberto Costa-Pazo, Sushil Bhattacharjee, Esteban Vazquez-Fernandez, and Sébastien Marcel. The replay-mobile face presentation-attack database. In *International Conference on Biometrics Special Interest Group (BIOSIG)*, 2016.

[11] Neslihan Erdogmus and Sébastien Marcel. Spoofing attacks to 2d face recognition systems with 3d masks. In *IEEE International Conference of the Biometrics Special Interest Group*, 2013.

[12] Javier Galbally, Sébastien Marcel, and Julián Fiérrez. Biometric antispoofing methods: a survey in face recognition. *IEEE Access*, 2:1530–1552, 2014.

[13] Javier Galbally, Sébastien Marcel, and Julian Fierrez. Image quality assessment for fake biometric detection: application to iris, fingerprint, and face recognition. *IEEE Transactions on Image Processing*, 23(2):710–724, 2014.

[14] JavaTpoint. Machine learning - support vector machine algorithm, Year.

[15] Firstname Lastname and Firstname Lastname. Presentation attack detection in face biometric systems using raw sensor data from smartphones. *Journal Name*.

[16] Xiaobai Li, Jukka Komulainen, Guoying Zhao, Pong C Yuen, and Matti Pietikäinen. Generalized face anti-spoofing by detecting pulse from face videos. In *International Conference on Pattern Recognition (ICPR)*, 2016.

[17] Amir H Mohammadi, Sushil Bhattacharjee, and Sébastien Marcel. Deeply vulnerable–a study of the robustness of face recognition to presentation attacks. *IET Biometrics*, 7(1):15–26, 2018.

[18] Keyurkumar Patel, Hu Han, and Anil K. Jain. Secure face unlock: Spoof detection on smartphones. *IEEE Transactions on Information Forensics and Security*, 11(10):2268–2283, 2016.

[19] R Raghavendra, Kiran B Raja, and Christoph Busch. Presentation attack detection for face recognition using light field camera. *IEEE Transactions on Image Processing*, 24(3):1060–1075, 2015.

[20] Daniel Raguin, George Mcclurg, Aleksei Sebastiani, Markus Schiefele, and Gregory Cannon. Presentation attack detection, 2020.

[21] Kiran B. Raja, Pankaj Wasnik, R. Raghavendra, and Christoph Busch. Robust face presentation attack detection on smartphones : An approach based on variable focus. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 651–658, 2017.

[22] Touseef Siddiqui, Soumyadeep Bharadwaj, Tejas Dhamecha, Ankan Agarwal, Mayank Vatsa, Richa Singh, and Nalini Ratha. Face anti-spoofing with multifeature videolet aggregation. In *International Conference on Pattern Recognition (ICPR)*, 2016.

[23] Jinyi Yang, Zhen Lei, Shengcai Liao, and Stan Z Li. Face liveness detection with component dependent descriptor. In *International Conference on Biometrics (ICB)*, 2013.