

Evaluation of Commercial Mobile Forensic Tools for Android Evidence Extraction in Digital Forensic Investigations

Aman Sharma^{1*}, Bhuvnesh Yadav¹, Bhavya Sharma²

¹Department of Chemistry, Biochemistry and Forensic Science, ASAS, Amity University Haryana, Haryana, India, ²Voice Division, Directorate of Forensic Services, Shimla Hills, Junga, Himachal Pradesh, India. *Corresponding Author's Email: amanforensic11@gmail.com

Abstract

Android smartphones are the major source of digital evidence in forensic investigations. However, different device models and operating systems make reliable data extraction challenging during forensic examination. This study evaluates comparative end-to-end artifact recovery from three commercial mobile forensic tools across different Android devices. Twenty Android smartphones of different makes and models were examined using Cellebrite UFED, MSAB XRY and Oxygen Forensic Detective. For each device-tool combination, the highest supported acquisition method available at the time of examination was used. The extracted artifacts were classified as call logs, contacts, chats, emails and multimedia. Tool performance was evaluated using dominance frequency analysis and pairwise Wilcoxon Signed-Rank Tests. Oxygen Forensic Detective achieved the highest call log and chat artifact counts in 16 and 17 of the 20 devices, respectively. Cellebrite UFED extracted the highest contact count in 12 of the devices, whereas MSAB XRY achieved the highest number of emails and multimedia files on 17 and 16 devices, respectively. Pairwise Wilcoxon testing identified statistically significant differences in several tool comparisons, although not all comparisons were significant. No single forensic tool consistently achieved the highest artifact recovery across all categories. The findings suggest that Android forensic extraction is device- and artifact-dependent. The count-based dominance should be interpreted as end-to-end recovery performance rather than as direct evidence of forensic accuracy or completeness.

Keywords: Android, Digital Evidence, Evidence Extraction, Forensic Tools, Mobile Forensics.

Introduction

The digital domain has evolved from a luxury to a necessity for individuals across generations, with smartphones becoming ubiquitous (1). Initially, mobile devices were introduced for voice communication and text messaging. However, their functionality has significantly expanded by serving as a source of entertainment, a means of capturing moments and facilitating easy communication, making smartphones an essential component of modern life (2). As technology has advanced, the chances of finding digital devices as evidence have also increased. Concurrently, digital forensics has expanded rapidly in recent years (3, 4). Digital devices generate vast amounts of information related to crimes, necessitating advanced analytical processes to keep pace with technological advancements (5). The wide range and types of digital crime underscore the need for evolved investigative procedures (6). Semiconductor technology enhances the computing power of mobile phones by transforming them into compact devices (7). The market encompasses a diverse

range of mobile devices with varying makes, models, capacities, properties and chipsets (8, 9). Generally, smartphones incorporate RAM (Random Access Memory), which is responsible for all executive functions, while NAND or NOR memory stores the operating system (10). Nearly every smartphone company uses an operating system such as Android, iOS, BlackBerry OS, Symbian, or Windows Phone (11). Undoubtedly, Google's Android and Apple's iOS are the dominant operating systems in the market (12). Android is an open-source Linux operating system launched by Google in 2005 (13). Google developed the Android OS to run powerful applications and provide users with convenient and user-friendly functionalities (11, 14). The Android architecture comprises five major layers (12, 15, 16). The application layer contains system and user-installed applications that interact with users and make use of services provided by the Android application framework. The second layer is the application framework, which includes components such as a

This is an Open Access article distributed under the terms of the Creative Commons Attribution CC BY license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution and reproduction in any medium, provided the original work is properly cited.

(Received 18th July 2026; Accepted 01st October 2026; Published 25th October 2026)

package manager, location manager, content provider and notification manager. The third and fourth layers are libraries and the Android runtime. Libraries consist of components written in C/C++ languages, such as SGL and SSL, whereas the Android runtime has two components: the Dalvik Virtual Machine and the Core Library. The lowest layer of Android architecture is the Linux kernel, which provides abstraction between hardware and software components, enables multitasking, manages storage and performs other such processes. The layered Android architecture and its interactive nature make forensic examination difficult (17). Forensic acquisition methods such as manual extraction, logical acquisition and hex-dump are discussed in several research studies, along with advanced techniques such as chip-off and JTAG (18).

In traditional digital forensics, investigation experts could quickly recover data due to the low storage capacity of devices (3). The period from 1999 to 2007 is considered the Golden Era of digital forensics, during which the origins of data recovery, including live and deleted data, were discovered (19). Research studies have established methods for forensic data extraction from smartphones that are scientifically valid for justice and law purposes (20, 21). Most of the mobile forensics research has focused on reviewing extraction techniques (22, 23). A mobile forensic foundation was formed with AT command data recovery methods (24). AT command protocol can access GSM devices, highlighting the significance of internal mobile memory over SIM or removable storage for data acquisition (25). Mobile forensic examination framework includes evidence identification, isolation, processing and archiving (26). Digital forensics is a scientific process that emphasizes the need for tool analysis and protocol validation (27). A hyper-formalized process that is often adopted in digital forensics leads to the development of scalable and tool-based analysis. Mobile operating system (OS) architectures present unique forensic challenges (28). Real-time investigation requires constant revision of protocols to deal with rapidly evolving device technologies and operating systems (26). The performance of extraction using Andriller, following the NIST forensic model, for BlackBerry Messenger data on Android devices demonstrated

the feasibility of a structured forensic framework in mobile application investigation (29).

Previous studies have shown that a mobile device may yield different extraction results when processed by different forensic tools (30, 31). The differences can be attributed to differences in device model, operating system version, security mechanism, acquisition support and artifact parsing capabilities (32). However, there is limited comparative evidence for the artifact-level performance of popular commercial forensic tools on heterogeneous Android devices from different manufacturers, versions of the operating system, chipsets and security patch levels. This is an important research gap, as a tool that performs well for one device or artifact category may not necessarily provide similar recovery for another. The novelty of the current study includes the comparative analysis of three commercial mobile forensic tools, namely Cellebrite UFED, MSAB XRY and Oxygen Forensic Detective, on 20 Android smartphones and five major artifact types: call logs, contacts, chats, emails and multimedia. In addition to the device-level comparison, the study also investigates the tool-dependent differences by applying dominance frequency analysis and pairwise Wilcoxon Signed-Rank Tests. The study thus provides comparative evidence at the level of devices and artifacts regarding recovery performance and demonstrates the practical value of the complementary multi-tool examination strategy. Accordingly, this study compares the artifact recovery performance of Cellebrite UFED, MSAB XRY and Oxygen Forensic Detective on different Android smartphones; investigates whether the extraction performance differs by device and artifact category; statistically compares paired differences in artifact recovery between the three forensic tools by employing the Wilcoxon Signed-Rank Test; and evaluates the practical value of an artifact-aware, multi-tool approach for Android forensic examination.

Methodology

Study Design

This experimental comparative study was conducted on Android smartphones using three globally used commercial mobile forensic tools. The study was conducted using standardised forensic acquisition methods under controlled laboratory conditions. The main aim of the study

was to evaluate artifact recovery from different forensic tools for different Android smartphones.

Device Selection

The study involved twenty Android smartphones from different device models and manufacturers, such as Samsung, Oppo, Vivo, Xiaomi and Realme. The devices were selected based on availability while maximising heterogeneity in manufacturer, model, Android version, chipset and security patch level. Inclusion criteria for the devices were regular use, full functionality and no known hardware defects at the time of examination. Devices not in working order or with hardware defects that may impede forensic acquisition were excluded. The

selection was made to have a heterogeneous set of Android devices with different manufacturers, models, operating system versions, chipsets and security patch levels. Table 1 summarizes the manufacturers, model numbers, Android operating system versions, chipsets and security patch levels of the 20 smartphones included in the study. The devices represent five manufacturers and a range of Android versions and chipset families. The security patch level recorded for each device represents the configuration present at the time of forensic examination and provides additional context for interpreting device-level extraction differences.

Table 1: Details of Mobile Phones Selected for Data Extraction

Make	Model	OS Version	Chipset	Patch Level
Samsung	SM-A305F	Android 11	Exynos 7904	2023-01-01
Samsung	SM-A127F	Android 13	Exynos 850	2023-12-01
Samsung	SM-A135F	Android 13	Exynos 850	2023-10-01
Samsung	SM-S908E	Android 13	Qualcomm SM8450 Snapdragon 8	2023-10-01
Samsung	SM-A146B	Android 14	Exynos 1330	2023-12-01
Samsung	SM-A336E	Android 14	Exynos 1280	2024-01-01
Samsung	SM-A546E	Android 14	Exynos 1380	2024-02-01
Samsung	SM-F926B	Android 14	Qualcomm SM8350 Snapdragon 888 5G	2024-01-01
Samsung	SM-G990E	Android 14	Qualcomm SM8350 Snapdragon 888 5G	2024-02-01
Vivo	V2151	Android 13	Qualcomm SM7325 Snapdragon 778G 5G	2023-08-01
Vivo	V2250	Android 13	Qualcomm SM7325 Snapdragon 778G 5G	2024-01-01
Xiaomi	Redmi Note 5	Android 8.1	Qualcomm SDM636 Snapdragon 636	2024-10-01
Xiaomi	Mi A1	Android 9	Qualcomm MSM8953 Snapdragon 625	2023-08-01
Oppo	CPH2185	Android 10	Mediatek MT6765	2023-02-05
Oppo	CPH1911	Android 11	Mediatek MT6771 Helio P70	2024-06-21
Oppo	CPH2213	Android 13	MediaTek Dimensity 800U	2024-11-05
Xiaomi	M2006C3LII	Android 10	MediaTek MT6762G	2024-11-02
Realme	RMX2189	Android 11	MediaTek MT6765G Helio G35	2023-07-05
Realme	RMX2117	Android 12	MediaTek Dimensity 800 U	2024-10-05
Realme	RMX3771	Android 13	MediaTek Dimensity 800 U	2024-12-05

Note: OS = Operating System, OEM = Original Equipment Manufacturer, MTK = MediaTek. The listed patch level represents the security patch level recorded for each device at the time of examination.

Commercial Extraction Tools

This study compares three widely used commercial mobile forensic tools: Cellebrite UFED 4PC (Version 7.71.181), MSAB XRY (Version 10.12.1) and Oxygen Forensic Detective (Version 18.1). These tools were selected as they are commercial-off-the-shelf platforms widely used for mobile data acquisition and analysis in digital forensic investigations. The selected tools are commercially available forensic platforms used in global digital forensic investigations. To maintain consistency, the same software version of each forensic tool was used throughout the study.

Device Preparation and Extraction

The SIM card and external memory card, if present, were removed before examination. All devices were fully charged, turned on and set to aeroplane mode. Developer options and USB debugging were enabled only when necessary for the supported acquisition workflow. Enabling OEM unlocking enabled the option to unlock the bootloader, but the bootloader was not unlocked unless required by the documented extraction method. The devices were examined using Cellebrite UFED, MSAB XRY and Oxygen Forensic Detective. The extracted records were separately compiled for each device, forensic tool and artifact category. Missing

extraction results were recorded as missing observations, not zero recovered artifacts. The recorded artifact counts were then used for dominance frequency analysis and for paired statistical comparison. Table 2 summarizes the

acquisition methods used by each forensic platform for the Android devices. The supported acquisition mode varied according to device model, Android version and tool capabilities.

Table 2: Extraction Methods Performed Using the Evaluated Forensic Tools

Device Model	Cellebrite UFED	MSAB XRY	Oxygen Forensic Detective
SM-A305F	File System	Logical	File System
SM-A127F	File System	Android Generic	Android Backup
SM-A135F	File System	File System	File System
SM-S908E	File System	Logical	Android Backup
SM-A146B	File System	Logical	Full File System
SM-A336E	File System	Logical	Android Backup
SM-A546E	File System	File System	Android Backup
SM-F926B	File System	Logical	Android Backup
SM-G990E	File System	File System	File System
V2151	File System	File System	File System
V2250	File System	Logical	File System
Redmi Note 5	File System	Android Generic	Android Backup
Mi A1	Physical	Physical	Physical
CPH2185	File System	File System	File System
CPH1911	File System	File System	File System
CPH2213	File System	Physical	Physical
M2006C3LII	File System	Logical	Physical
RMX2189	File System	Android Generic	Android Backup
RMX2117	File System	Logical	File System
RMX3771	File System	Android Backup	Android Backup

Note: UFED = Universal Forensic Extraction Device, XRY = XRY Forensic Extraction Platform. “-” indicates that no extraction result was available for the corresponding device-tool combination.

The acquired datasets were then parsed and analyzed with the respective forensic analysis software. The extracted records were divided into five types of artifacts: call logs, contacts, chats, emails and multimedia. For each device, the number of artifacts reported by each forensic platform was recorded. The acquisition level supported did not match for all device-tool combinations, so the results are a comparison of end-to-end forensic tool performance, not a comparison of performance using the same source images.

Statistical Analysis

For the statistical analysis, dominance frequency analysis and Wilcoxon Signed-Rank Test were performed. Data were analyzed using IBM SPSS Statistics for Windows, Version 32.0 (IBM Corp., Armonk, NY, USA). Dominance Frequency Analysis was used to compare tool performance across all artifact categories. For each device model, the forensic tool that yielded the highest number of extracted artifacts was considered dominant for

that specific artifact. Dominance frequencies were then calculated across all devices on an artifact-wise basis. This analysis provides a device-level summary of how frequently each forensic tool produced the highest artifact count within each artifact category, without treating differences in absolute artifact counts as direct measures of forensic accuracy.

The Wilcoxon Signed-Rank Test was selected because of the paired structure of the data and non-normal distribution of the differences. It was performed to determine the significance of extracted artifact data from each tool. The Wilcoxon Signed-Rank Test was chosen because the differences in device-level artifact count pairs could not be assumed to be normally distributed and the sample size ($n=20$ devices) was relatively small. Missing extraction observations, represented as ‘-’ in the figures, were treated as missing values and excluded pairwise from the corresponding Wilcoxon Signed-Rank Test; they were not interpreted as zero retrieved artifacts.

Results

The dataset extracted using Cellebrite UFED, MSAB XRY and Oxygen Forensic Detective was classified into five primary artifact categories: call logs, contacts, chats (including SMS and social media messaging applications), emails and multimedia data (images, videos, audio files, documents). The results revealed considerable variation in extraction outcomes, strongly influenced by both device model and artifact type:

Artifact-wise Extraction Performance

To conclude the technical comparison, artifact-wise discrepancies in data extraction represent varying degrees of complexity.

Call Logs: The call-log recovery showed comparatively lower variability than other artifact categories. Figure 1 illustrates the device-wise variation in call-log extraction among Cellebrite UFED, MSAB XRY and Oxygen Forensic Detective. The extraction counts varied across both devices and forensic tools, indicating that call-log recovery was influenced by the specific device-tool combination. Samsung devices exhibited diverse extraction behaviour. The device SM-F926B recorded the highest call log recovery across all tools. Oxygen Forensic Detective, however, failed to extract call logs from SM-G990E, whereas UFED and XRY successfully retrieved records, indicating

a tool-specific limitation. Oxygen Forensic Detective produced substantially higher counts than UFED and XRY on SM-S908E, SM-A127F and SM-A305F. In contrast, XRY extracted the highest number of call logs from SM-A135F. Oppo devices (CPH2185, CPH1911 and CPH2213), Oxygen Forensic Detective consistently extracted a higher number of call logs than UFED and XRY. For the Vivo V2151, UFED and XRY showed moderate recovery, while Oxygen Forensic Detective extracted a noticeably high number of call logs. Among Xiaomi devices, UFED and Oxygen Forensic Detective performed comparably on Mi A1, whereas XRY extracted the highest number of call logs from Redmi Note 5. Across Realme devices (RMX3771, RMX2117, RMX2189), Oxygen Forensic Detective produced higher artifact counts than the other tools.

Oxygen Forensic Detective achieved the highest call-log count in 16 of the 20 devices, while MSAB XRY achieved the highest count in 3 devices. The only device with a tie was the Xiaomi Mi A1. Both Cellebrite UFED and Oxygen Forensic Detective recovered 512 call-log records. Cellebrite UFED did not achieve dominance for call-log extraction on any device; however, one tie between UFED and Oxygen Forensic Detective occurred for the device model Xiaomi Mi A1.

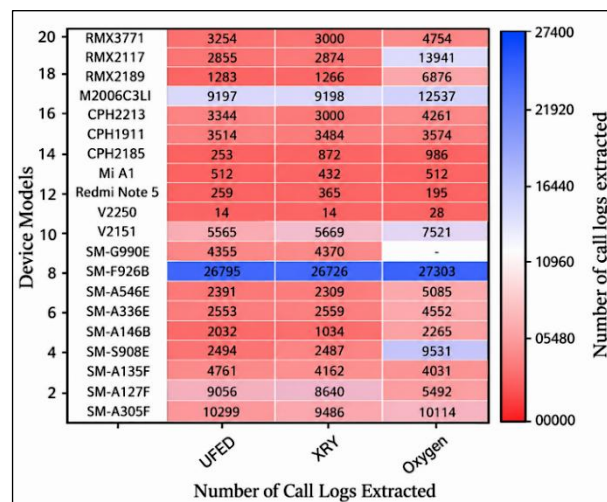


Figure 1: Heatmap Representation of Call Logs Extracted from Devices Using Cellebrite UFED, MSAB XRY and Oxygen Forensic Detective

Note: UFED = Universal Forensic Extraction Device, XRY = XRY Forensic Extraction Platform, - = no extraction results available.

Contacts: The extraction of contacts varied considerably across all devices and tools. Figure 2 shows the device-wise distribution of contact artifacts recovered by the three forensic tools. Among Samsung devices, UFED and XRY generally

extracted a higher volume of contacts. Oxygen Forensic Detective failed to recover contact data from SM-G990E, while UFED and XRY extracted moderate data. UFED consistently outperformed other tools in SM-A127F, SM-A305F and SM-

A135F, whereas XRY showed marginal superiority in SM-F926B. In Oppo devices, UFED demonstrated the strongest performance in CPH1911 and CPH2213. Notably, XRY failed to extract contacts from CPH2185, while UFED and Oxygen Forensic Detective succeeded. Vivo devices displayed mixed behaviour: XRY dominated in V2151, whereas Oxygen Forensic Detective extracted the highest number of contacts from V2250. For Xiaomi devices, overall contact recovery was comparatively low. XRY failed to extract contacts from Mi A1, while UFED and Oxygen Forensic Detective succeeded. In the Redmi Note 5, XRY achieved the highest recovery.

Among Realme devices, UFED outperformed other tools in RMX3771 and RMX2117, whereas Oxygen Forensic Detective extracted the highest number of contacts from RMX2189.

Overall, UFED demonstrated the most consistent contact extraction across manufacturers, particularly Samsung and Oppo. XRY showed strong performance in selected devices but exhibited complete extraction failures in other devices. Oxygen Forensic Detective achieved higher recovery in certain Vivo, Realme and Xiaomi devices but lacked consistency across tested Samsung models.

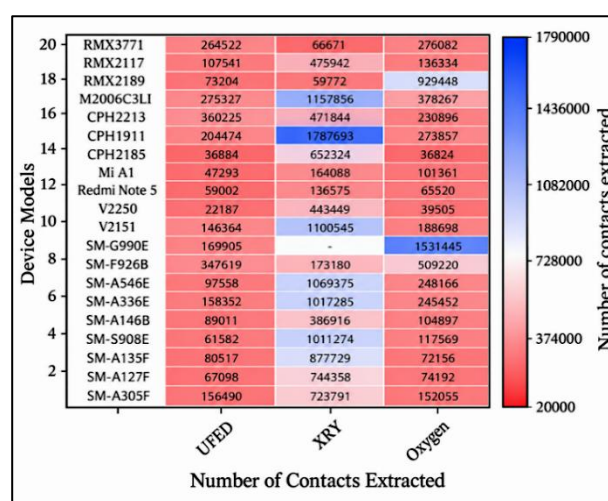


Figure 2: Heatmap Representation of Contact Extracted from Devices Using Cellebrite UFED, MSAB XRY and Oxygen Forensic Detective

Note: UFED = Universal Forensic Extraction Device, XRY = XRY Forensic Extraction Platform, - = no extraction results available.

Chats (SMS and Application Data): Chat data extraction, including SMS and third-party messaging applications, showed the greatest variability and tool dependency. Figure 3 presents the comparative extraction of chat artifacts. In Samsung devices, Oxygen Forensic Detective extracted substantially higher chat data in most models, including SM-F926B, SM-A336E, SM-A546E, SM-A127F and SM-A305F. However, in SM-G990E, Oxygen Forensic Detective failed, while XRY extracted the highest number of chat records. SM-A146B further highlighted tool inconsistency, where XRY failed, but Oxygen Forensic Detective recovered a large volume of chat data. Across Oppo devices, Oxygen Forensic Detective consistently outperformed UFED and XRY, with XRY failing in CPH2185. Vivo devices showed moderate chat

extraction using UFED and XRY, whereas Oxygen Forensic Detective consistently recovered higher volumes. In Xiaomi devices, Oxygen Forensic Detective dominated in Mi A1, while XRY extracted the highest number of chats from Redmi Note 5. For all tested Realme devices, Oxygen Forensic Detective consistently extracted significantly higher chat records than UFED and XRY.

Oxygen Forensic Detective showed the highest number of chat artifacts on 17 of the 20 devices tested, showing the greatest dominance in this artifact category. MSAB XRY achieved the highest chat-artifact count in three devices: SM-S908E, SM-G990E and Redmi Note 5. Cellebrite UFED did not achieve the highest chat-artifact count for any of the devices evaluated.

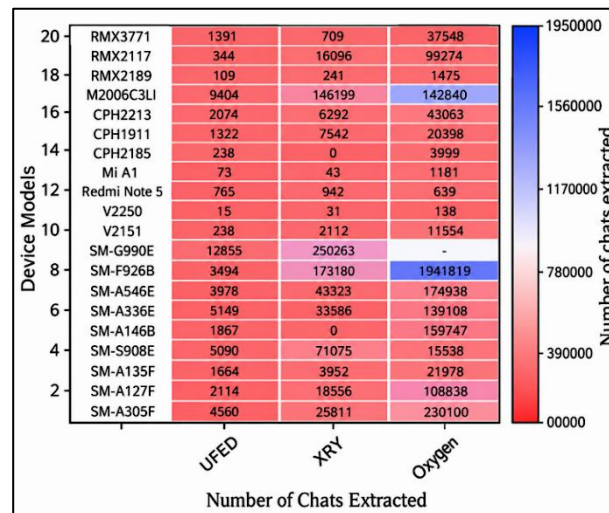


Figure 3: Heatmap Representation of Chats Extracted from Devices Using Cellebrite UFED, MSAB XRY and Oxygen Forensic Detective

Note: UFED = Universal Forensic Extraction Device, XRY = XRY Forensic Extraction Platform, - = no extraction results available.

Emails: Email artifacts extracted from applications such as Gmail, Yahoo Mail and other mail apps revealed comparatively consistent patterns across devices. Figure 4 illustrates the comparative recovery of email artifacts from applications. In Samsung devices, XRY outperformed UFED and Oxygen Forensic Detective in nearly all models. Oxygen Forensic Detective failed to extract emails from SM-G990E, while XRY recorded the highest recovery. Similar dominance by XRY was observed in SM-F926B, SM-A336E, SM-A546E, SM-A135F, SM-A127F and SM-A305F. Among Oppo and Vivo devices, XRY consistently achieved the highest email recovery. In Xiaomi devices, XRY dominated in Mi A1, whereas Oxygen Forensic Detective extracted more emails from Redmi Note 5. For

Realme devices, XRY outperformed other tools in RMX2117 and RMX2189, while Oxygen Forensic Detective demonstrated higher extraction in RMX3771. MSAB XRY demonstrated the highest dominance frequency for email artifacts, successfully dominating 17 out of 20 devices. Oxygen Forensic Detective dominated in 3 devices, while Cellebrite UFED showed minimal dominance. The pattern observed shows relatively consistent email artifact recovery by MSAB XRY across the tested devices; however, the present study did not independently assess the technical reasons for these differences. Compared with chat artifacts, email extraction exhibits lower variability, although tool-dependent differences persist.

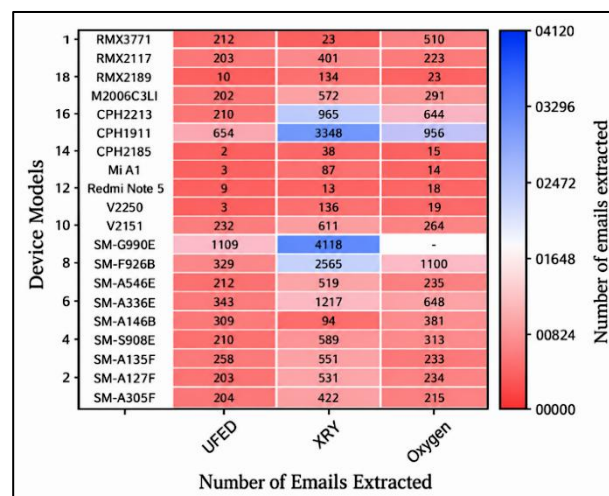


Figure 4: Heatmap Representation of Emails Extracted from Devices Using Cellebrite UFED, MSAB XRY and Oxygen Forensic Detective

Note: UFED = Universal Forensic Extraction Device, XRY = XRY Forensic Extraction Platform, - = no extraction results available.

Multimedia: Multimedia extraction results showed extreme variability across devices and tools. Figure 5 presents the device-wise comparison of multimedia artifacts recovered using the three forensic tools. In Samsung devices, Oxygen Forensic Detective extracted the highest multimedia volume from SM-G990E and SM-F926B, while XRY failed in SM-G990E. Conversely, XRY extracted higher multimedia volumes in several mid-range Samsung models. For Oppo, Vivo and Xiaomi devices, XRY consistently extracted the highest number of multimedia records. In Realme devices, XRY dominated in RMX2117; Oxygen Forensic Detective extracted substantially higher multimedia data from RMX2189 and UFED and Oxygen Forensic

Detective showed comparable performance in RMX3771.

The extraction of multimedia artifacts showed a notable variability among the forensic tools investigated. Most multimedia artifacts were detected by MSAB XRY on 16 of 20 devices, particularly on Samsung, Oppo, Vivo and Xiaomi models. Oxygen Forensic Detective had the highest count on 4 devices: SM-F926B, SM-G990E, RMX2189 and RMX3771. For all the evaluated devices, Cellebrite UFED did not achieve the highest multimedia artifact count. The results show that the multimedia recovery was highly dependent on the forensic tool and device-specific acquisition support.

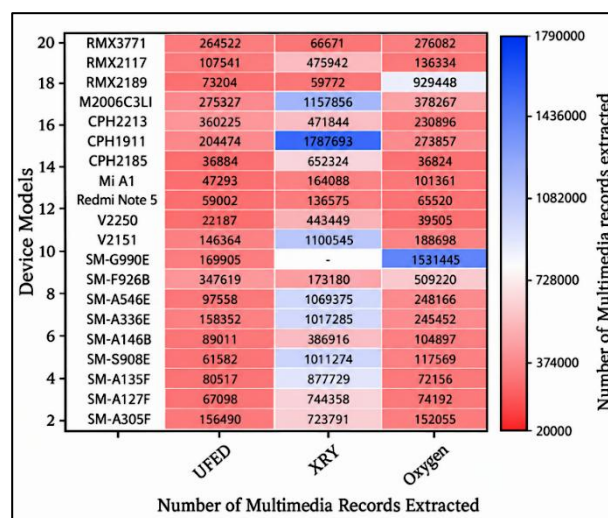


Figure 5: Heatmap Representation of Multimedia Extracted from Devices Using Cellebrite UFED, MSAB XRY and Oxygen Forensic Detective.

Note: UFED = Universal Forensic Extraction Device, XRY = XRY Forensic Extraction Platform, - = no extraction results available.

Chipset-based Analysis

The devices were further analyzed according to their chipsets. The chipsets were categorized into Samsung Exynos, Qualcomm Snapdragon and MediaTek (MTK) chips. The Samsung devices with Exynos chips showed greater differences primarily in chats and multimedia artifacts. The Oxygen Forensic Detective demonstrated higher recovery of chat artifacts, while MSAB XRY performed better for email artifacts for Exynos chipsets. For Qualcomm Snapdragon chipsets, smaller differences in extraction were observed for call logs and contact artifacts. However, application-based artifacts showed variability during analysis.

Statistical Findings

Dominance Frequency Analysis: For each artifact category, the dominating extractor for each device was identified separately. Ties were recorded separately and not assigned to any tool. The tie for call-log extraction for the Xiaomi Mi A1 was 512 records recovered by both Cellebrite UFED and Oxygen Forensic Detective. Missing extraction results were treated as missing observations, not as zero artifacts recovered. In general, the highest dominance frequency for call logs and chats was shown by Oxygen Forensic Detective, for contacts – by Cellebrite UFED and for emails and multimedia – by MSAB XRY. The complete dominance-frequency distribution is presented in Table 3.

Table 3: Dominance Frequencies Across the Five Artifact Categories

Artifact category	Dominance Frequency	Oxygen Forensic Detective	UFED	XRY
Call Logs	n = 20 devices	16	0	3
Contact	n = 20 devices	2	12	6
Chats	n = 20 devices	17	0	3
Emails	n = 20 devices	3	0	17
Multimedia	n = 20 devices	4	0	16

Note: UFED = Universal Forensic Extraction Device, XRY = XRY Forensic Extraction Platform, – = no extraction results available. Dominance is defined as the tool that returns the highest number of artifacts for a given device and should not be interpreted as evidence of superior accuracy or completeness on its own. For call logs, one device (Xiaomi Mi A1) produced a tie between UFED and Oxygen Forensic Detective and was therefore not assigned to either tool.

Pairwise Wilcoxon Signed-rank Test

Pairwise Wilcoxon Signed-Rank Tests revealed statistical differences depending on the artifact category and tool comparison. For call logs, the statistically significant difference was not found between UFED and XRY ($p=0.780$), but a significant difference was found between UFED and Oxygen Forensic Detective ($p=0.0007$) and between XRY and Oxygen Forensic Detective ($p=0.004$). For contact, there were statistically significant differences between UFED and XRY ($p = 0.012$) and between XRY and Oxygen Forensic Detective ($p = 0.039$) but not between UFED and Oxygen Forensic Detective ($p = 0.430$). For chat artifacts, all three pairwise comparisons were significant: UFED vs. XRY ($p = 0.036$), UFED vs. Oxygen Forensic

Detective ($p = 0.000006$) and XRY vs. Oxygen Forensic Detective ($p = 0.000483$). All three email artifact comparisons were also statistically significant (UFED vs. XRY, $p = 0.000013$; UFED vs. Oxygen Forensic Detective, $p = 0.000708$; XRY vs. Oxygen Forensic Detective, $p = 0.000586$). There were statistically significant differences between UFED and XRY ($p = 0.000708$) and UFED and Oxygen Forensic Detective ($p = 0.000002$) for multimedia artifacts, but not between XRY and Oxygen Forensic Detective ($p = 0.153646$). The complete W statistics and corresponding p-values are presented in Table 4. P-values were not corrected for multiple comparisons and these statistical results are to be interpreted as exploratory.

Table 4: Pairwise Wilcoxon Signed-Rank Test Results for Artifact Extraction

Artifact	Comparison	W Statistic	p-value
Call Logs	UFED vs XRY	68	0.78
	UFED vs Oxygen	14	0.0007
	XRY vs Oxygen	18	0.004
Contact	UFED vs XRY	36	0.012
	UFED vs Oxygen	42	0.430
	XRY vs Oxygen	72	0.039
Chats	UFED vs XRY	78	0.036
	UFED vs Oxygen	0	0.000006
	XRY vs Oxygen	0	0.000483
Emails	UFED vs XRY	3	0.000013
	UFED vs Oxygen	62	0.000708
	XRY vs Oxygen	6	0.000586
Multimedia	UFED vs XRY	8	0.000708
	UFED vs Oxygen	22	0.000002
	XRY vs Oxygen	58	0.153646

Note: UFED = Universal Forensic Extraction Device, XRY = XRY Forensic Extraction Platform. W represents the Wilcoxon Signed-Rank Test statistic. All statistical tests were two-tailed. The reported p-values were not adjusted for multiple comparisons; therefore, the results should be interpreted as exploratory.

The results indicate that no single forensic tool yielded the highest number of artifacts across all categories. Oxygen Forensic Detective had the highest count-based dominance for call logs and chats, Cellebrite UFED for contacts and MSAB XRY for emails and multimedia. The Wilcoxon Signed-

Rank results also indicated that both artifact category and tool combination affected pairwise differences associated with differences between paired observations. Critically, dominance frequency and Wilcoxon testing assess distinct dimensions of comparative performance:

dominance determines which tool most frequently achieved the highest count, while the Wilcoxon test examines paired device-level disparities. Reported p-values were not adjusted for multiple comparisons and therefore statistically significant findings should be considered exploratory.

Discussion

In the present study, artifact recovery using commercial mobile forensic tools was device and artifact-dependent. No single tool dominated across all five categories. Oxygen Forensic Detective showed the highest count-based dominance for call logs and chats, Cellebrite UFED in contact and MSAB XRY in emails and multimedia. These observed differences may reflect variations in acquisition support, Android and OEM implementations, security mechanisms, application storage structures and artifact-parsing capabilities (33, 34).

For call logs, Oxygen Forensic Detective had the highest count in 16 of the 20 devices, XRY had the highest count in three devices and there was one tie between UFED and Oxygen. The Wilcoxon analysis showed no statistically significant differences between UFED and XRY. Statistically significant differences were found between UFED and Oxygen Forensic Detective and between XRY and Oxygen Forensic Detective. Oxygen frequently achieved higher call-log counts across evaluated devices, but performance was found to be device-dependent.

Contact artifacts showed a contrasting trend, with Cellebrite UFED achieving the highest contact-artifact count in a greater number of the evaluated devices. However, the present study did not independently examine the database structure, encryption status, or parser-level mechanisms responsible for these differences. Therefore, the observed variations should primarily be interpreted as differences in end-to-end artifact recovery among the evaluated forensic platforms. Previous studies have also reported tool-dependent differences in the recovery of information-management artifacts using commercial mobile forensic platforms (35, 36). These findings also highlight challenges introduced by encryption, sandboxing and scoped storage for devices with newer Android versions. Chat artifacts showed strong tool dependence. Oxygen Forensic Detective had the highest chat-

artifact count in 17 devices, XRY in three devices and UFED did not lead on any device. All three pairwise Wilcoxon comparisons were statistically significant at the unadjusted 0.05 level. The observed variability may be associated with the complexity of extracting application-based communications from Android devices, where factors such as encryption, sandboxing, application version, acquisition method and tool-specific parsing support can influence artifact availability (37, 38).

Email extraction revealed a different pattern. The MSAB XRY tool produced the greatest number of email artifacts in 17 of the 20 devices tested. All three tool comparisons were also statistically significantly different for email artifacts using pairwise Wilcoxon Signed-Rank Testing. However, dominance frequency and Wilcoxon testing measure different aspects of comparative performance and should not be interpreted interchangeably. The variation observed may be due to differences in application storage, acquisition support, synchronisation mechanisms and artifact parsing, but the present study did not independently evaluate the technical mechanisms responsible for these differences (39). The chat and email artifact performance underscores a central outcome that the effectiveness of forensic tools is tightly coupled with artifact types. The pattern of multimedia artifacts was different. MSAB XRY produced the highest multimedia-artifact count in 16 of the 20 devices and Oxygen Forensic Detective the highest in four devices. UFED did not achieve count-based dominance in this artifact category. The Wilcoxon comparison between XRY and Oxygen was not statistically significant ($p = 0.153646$). This is an example of why these two statistics should not be used interchangeably; a tool can have the highest count across more devices but not have enough paired differences to be statistically significant (40, 41).

Several studies compared the tool performance of commercial and open-source digital forensic tools. Second-hand devices were analyzed using UFED, MSAB XRY and RTA, resulting in variability of recovered artifacts among these tools (30). Comparison of MOBILedit, UFED, SANS SIFT and the Sleuth Kit demonstrated that commercial tools usually excelled in speed and data recovery, whereas open-source tools are more cost-effective (31). In another research study, FTK Imager and

Paraben device seizure outperformed EnCase, MOBILedit and Oxygen Forensic Detective in data extraction (32, 42). The data extraction using SP Flash Tool, Cellebrite UFED and Volcano Box on MTK-based Android phones highlighted tool-specific strengths, with SP Flash Tool demonstrating superior compliance (43).

A significant methodological component of the present study was the employment of paired non-parametric analysis to compare the artifact counts at the device level across the evaluated forensic tools. The Wilcoxon Signed-Rank Test was appropriate for paired device-level data where normality of differences could not be assumed (41).

The results of the present study support the use of a multi-tool, artifact-aware approach to Android forensic examination. However, the number of artifacts reported by a forensic tool does not automatically mean forensic accuracy. Reported counts may vary due to differences in acquisition level, parser implementation, duplicate handling, cached data and device-specific support. Thus, the results of this investigation are a description of comparative end-to-end artifact recovery and not sensitivity, specificity, or error rate unless an independent ground-truth dataset is available. Complementary tools may thus be useful in operational investigations for cross-validation and for minimizing the risk of missing potentially relevant artifacts.

From a practical perspective, the findings indicate that forensic examiners should not depend on a single commercial forensic platform if a comprehensive Android artifact recovery is required. In addition to the artifact category of investigative interest, the device model, supported acquisition level and operating system characteristics should be considered in the tool selection process. Where operationally possible, the use of complementary forensic tools may assist with cross-validating recovered artifacts and identifying data that may not be parsed or recovered by another platform. Examiners should also be aware of the acquisition method and tool version involved in each examination, as variances in acquisition support may directly impact the artifacts available for further analysis. This research was limited to 20 Android smartphones and three commercial forensic tools. Not all combinations of device and tool used the same

acquisition method and no independent ground-truth dataset was used; thus, artifact counts should not be interpreted as direct measures of forensic accuracy or completeness. Future research should incorporate more Android devices, additional forensic tools and controlled ground-truth datasets to further evaluate the extraction performance.

Conclusion

This study evaluated the end-to-end artifact recovery performance of three commercial mobile forensic platforms (Cellebrite UFED, MSAB XRY and Oxygen Forensic Detective) on 20 heterogeneous Android smartphones with different manufacturers, operating-system versions, chipsets and security patch levels. Results indicated that there were substantial differences in artifact recovery across device models and artifact types. Oxygen Forensic Detective showed the highest number-based dominance for call logs and chats, Cellebrite UFED for contacts and MSAB XRY for emails and multimedia. Pairwise Wilcoxon signed-rank analysis further demonstrated that differences between forensic tools were artifact category and specific tool comparison dependent. The principal contribution of this research is an artifact-level and device-level comparison of three popular commercial mobile forensic platforms under controlled laboratory conditions. The observations indicated that forensic extraction performance cannot be adequately described by considering a forensic tool independent of the device and artifact under consideration. The practical implications are relevant for the forensic examination workflows where the recovery of mobile artifacts needs to be comprehensive. The results obtained indicate that use of a single forensic platform may not be sufficient to fully recover some categories of artifacts on some device configurations. Thus, complementary forensic platforms can be used, where operationally feasible, to cross-check recovered artifacts and identify discrepancies between extraction results. There are some important limitations to the interpretation of the findings. The investigation was limited to 20 Android smartphones and three commercial forensic tools and the acquisition method was dependent on the capabilities supported for each device-tool combination. These limitations make it difficult to

generalize the observed differences to all Android devices, operating-system versions, applications, or forensic environments. Future work should extend the device sample to include other manufacturers, Android versions, chipsets, security configurations and application versions. Future work should also include a controlled ground-truth dataset with known artifacts. Additional commercial and open-source forensic platforms should be evaluated in further studies and the effect of acquisition method, parser implementation, encryption, application version and duplicate handling on the artifact recovery should be studied.

Abbreviations

Android OS: Android operating system, JTAG: Joint Test Action Group, MTK: MediaTek, NIST: National Institute of Standards and Technology, OEM: Original Equipment Manufacturer, SMS: Short Message Service, UFED: Universal Forensic Extraction Device, XRY: XRY forensic extraction platform.

Acknowledgement

None

Author Contributions

Aman Sharma: Conceptualization, Study Design, Methodology, Investigation, Data Acquisition, Data Curation, Statistical Analysis, Interpretation of Results, Manuscript Preparation, Bhuvnesh Yadav: Conceptualization, Study Design, Methodology, Investigation, Data Acquisition, Data Curation, Statistical Analysis, Interpretation of Results, Manuscript Preparation, Bhavya Sharma: Conceptualization, Study Design, Methodology, Investigation, Data Acquisition, Data Curation, Statistical Analysis, Interpretation of Results, Manuscript Preparation. All authors contributed to reviewing and editing the manuscript, approved the final version and agreed to be accountable for the work.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability

The data generated and analyzed during this study are available from the corresponding author upon reasonable request. The dataset includes device-level extraction results obtained from the commercial mobile forensic tools evaluated in this

study. Any data shared will be subject to applicable forensic, institutional and confidentiality requirements.

Declaration of Artificial Intelligence (AI) Assistance

The authors acknowledge the use of generative AI to assist in the paraphrasing of certain parts of the manuscript. All content generated through AI was critically reviewed, edited and verified by the authors to ensure accuracy, originality and compliance with academic standards. The authors take full responsibility for the integrity and validity of the manuscript. The figures presented in this manuscript were generated from the authors' experimental data and were not generated using generative artificial intelligence.

Ethics Approval

Institutional Ethics Committee approval was not required, as the study involved exclusively investigator-owned laboratory devices and generated datasets and did not involve human participants or any identifiable personal information.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

1. Irons A, Lallie HS. Digital Forensics to Intelligent Forensics. *Future Internet*. 2014;6(3):584–596. doi:10.3390/fi6030584
2. Feng P, Li Q, Zhang P, *et al*. Logical acquisition method based on data migration for Android mobile devices. *Digit Investig*. 2018;26:55–62. doi:10.1016/j.diin.2018.05.003
3. Garfinkel SL. Digital forensics research: The next 10 years. *Digit Investig*. 2010;7:S64–S73. doi:10.1016/j.diin.2010.05.009
4. Gupta GK. Ubiquitous mobile phones are becoming indispensable. *ACM Inroads*. 2011;2(2):32–33. doi:10.1145/1963533.1963545
5. Nosál J. Crime in the digital age: a new frontier. In: Berghofer J, Futter A, Häusler C, Hoell M, Nosál J, editors. *The Implications of Emerging Technologies in the Euro-Atlantic Space*. Cham: Palgrave Macmillan. 2023;177–193. doi:10.1007/978-3-031-24673-9
6. Jones GM, Winster SG. An insight into digital forensics: history, frameworks, types and tools. *Cyber Security and Digital Forensics*. Wiley. 2022;105–125. doi:10.1002/9781119795667.ch6
7. Yeboah LA, Oppong PA, Malik A, *et al*. Exploring innovations, sustainability and future opportunities

- in semiconductor technologies. *Int J Adv Nano Comput Anal.* 2024;3(2).
doi:10.61797/ijanca.v3i2.348
8. Cecere G, Corrocher N, Battaglia RD. Innovation and competition in the smartphone industry: Is there a dominant design? *Telecomm Policy.* 2015;39(3-4):162-175.
doi:10.1016/j.telpol.2014.07.002
 9. Sun L, Fah CT. Xiaomi: transforming the competitive smartphone market to become a major player. *Eurasian J Soc Sci.* 2020;8(3):96-110.
doi:10.15604/ejss.2020.08.03.002
 10. Meena JS, Sze SM, Chand U, *et al.* Overview of emerging nonvolatile memory technologies. *Nanoscale Res Lett.* 2014;9(526).
doi:10.1186/1556-276X-9-526
 11. Sarkar A, Goyal A, Hicks D, *et al.* Android application development: a brief overview of Android platforms and evolution of security systems. 2019 Third International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). IEEE. 2019;73-79.
doi:10.1109/I-SMAC47947.2019.9032440.
 12. Al-Hadadi M, AlShidhani A. Smartphone forensics analysis: a case study. *IJCEE.* 2013;5(6):576-580.
doi:10.7763/IJCEE.2013.V5.776
 13. Khunt AR, Prabu P. An Empirical Analysis of Android Permission System Based on User Activities. *J Comput Sci.* 2018;14(3):324-333.
doi:10.3844/jcssp.2018.324.333
 14. Dar MA, Parvez J. Smartphone operating systems: evaluation & enhancements. 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT). IEEE. 2014;734-738.
doi:10.1109/ICCICCT.2014.6993056
 15. Mylonas A, Meletiadis V, Mitrou L, Gritzalis D. Smartphone sensor data as digital evidence. *Comput Secur.* 2013;38:51-75.
doi:10.1016/j.cose.2013.03.007
 16. Wei X, Gomez L, Neamtii I, Faloutsos M. ProfileDroid: multi-layer profiling of Android applications. *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking.* ACM. 2012;137-148.
doi:10.1145/2348543.2348563
 17. Padshah M, Gadakh SS, Sharma DM, Gongal NB. A Comparative Analysis of the Architectural Foundations of Android and iOS. 2025 5th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS). 2025;189-195.
doi:10.1109/ICUIS67429.2025.11380493
 18. Almuqren A, Alsuaelim H, Hafizur Rahman MM, Ibrahim AA. A Systematic Literature Review on Digital Forensic Investigation on Android Devices. *Procedia Comput Sci.* 2024;235:1332-1352.
doi:10.1016/j.procs.2024.04.126
 19. Nance K, Hay B, Bishop M. Digital forensics: defining a research agenda. In: 2009 42nd Hawaii International Conference on System Sciences. 2009;1-6.
doi:10.1109/HICSS.2009.160
 20. Barmapsalou K, Damopoulos D, Kambourakis G, Katos V. A Critical Review of 7 Years of Mobile Device Forensics. *Digit Investig.* 2013;10(4):323-349.
doi:10.1016/j.diin.2013.10.003
 21. Reedy P. Interpol review of digital evidence 2016-2019. *Forensic Sci Int Synerg.* 2020;2:489-520.
doi:10.1016/j.fsisyn.2020.01.015
 22. Chernyshev M, Zeadally S, Baig Z, Woodward A. Mobile forensics: advances, challenges and research opportunities. *IEEE Secur Priv.* 2017;15(6):42-51.
doi:10.1109/MSP.2017.4251107
 23. Barmapsalou K, Cruz T, Monteiro E, Simões P. Mobile Forensic Data Analysis: Suspicious Pattern Detection in Mobile Evidence. *IEEE Access.* 2018;6:59705-59727.
doi:10.1109/ACCESS.2018.2875068
 24. James JE. Foundations of mobile forensics: an academic approach. *Issues Inf Syst.* 2024;25(3):94-108.
doi:10.48009/3_iis_2024_108
 25. Figueiredo RC, Ribeiro AMO, Arthur R, Conforti E. Remote SMS instrumentation supervision and control using LabVIEW. *Practical Applications and Solutions Using LabVIEW Software.* InTech. 2011;317-340.
doi:10.5772/22540
 26. Islam S, Aziz B, Blackwell C. A Framework for Digital Forensics and Investigations. *Int J Digit Crime Forensics.* 2013;5(2):1-22.
doi:10.4018/jdcf.2013040101
 27. Beebe N. Digital forensic research: the good, the bad and the unaddressed. *Advances in Digital Forensics V. Berlin/Heidelberg: Springer.* 2009;17-36.
doi:10.1007/978-3-642-04155-6_2.
 28. Yusoff MN, Mahmood R, Abdullah MT, Dehghantanha A. Mobile Forensic Data Acquisition in Firefox OS. *SCISPACE.* 2014;27-31.
doi:10.1109/CyberSec.2014.6913967
 29. Riadi I, Umar R, Firdonsyah A. Forensic tools performance analysis on Android-based Blackberry Messenger using NIST measurements. *IJECE.* 2018;8(5):3991-4003.
doi:10.11591/ijece.v8i5.pp3991-4003
 30. Glisson WB, Storer T, Buchanan-Wollaston J. An Empirical Comparison of Data Recovered from Mobile Forensic Toolkits. *Digit Investig.* 2013;10(1):44-55.
doi:10.1016/j.diin.2013.03.004
 31. Padmanabhan R, Lobo K, Ghelani M, Sujan D, Shirole M. Comparative Analysis of Commercial and Open Source Mobile Device Forensic Tools. 2016 Ninth International Conference on Contemporary Computing (IC3). IEEE. 2016;1-6.
doi:10.1109/IC3.2016.7880238
 32. Osho O, Ohida SO. Comparative Evaluation of Mobile Forensic Tools. *IJITCS.* 2016;8(1):74-83.
doi:10.5815/ijitcs.2016.01.09
 33. Bandara V, Pletinckx S, Grishchenko I, Kruegel C, Vigna G, Tapiador J, Vallina-Rodriguez N. Beneath the Surface: An Analysis of OEM Customizations on the Android TLS Protocol Stack. 2025 IEEE 10th European Symposium on Security and Privacy (EuroS&P). 2025;677-700.
doi:10.1109/EuroSP63326.2025.00044
 34. Chen J, Li K, Chen Y, Wei L, Liu Y. Demystifying Device-Specific Compatibility Issues in Android Apps. 2024 IEEE International Conference on Software Maintenance and Evolution (ICSME). IEEE. 2024;525-537.

- doi:10.48550/arXiv.2408.01810
35. Lessard J, Kessler GC. Android forensics: simplifying cell phone examinations. *Small Scale Digit Device Forensics J.* 2010;4(1):1-12. ISSN:1941-6164.
https://www.garykessler.net/library/SSDDFJ_V4_1_Lessard_Kessler.pdf
 36. Hoog A. Android forensics: investigation, analysis and mobile security for Google Android. Elsevier. 2011;432.
https://books.google.co.in/books/about/Android_Forensics.html?id=zDibrpXTfxMC&redir_esc=y
 37. Mahajan A, Dahiya MS, Sanghvi HP. Forensic analysis of instant messenger applications on Android devices. *International Journal of Computer Applications.* 2013;68(8):38-44. doi:10.5120/11602-6965
 38. Vidas T, Zhang C, Christin N. Toward a general collection methodology for Android devices. *Digit Investig.* 2011;8:S14-S24. doi:10.1016/j.diin.2011.05.003
 39. Casey E. Digital evidence and computer crime: forensic science, computers and the Internet. 3rd ed. Waltham, MA: Academic Press. 2011;807.
 - https://booksite.elsevier.com/samplechapters/9780123742681/Front_Matter.pdf
 40. Al Mutawa N, Baggili I, Marrington A. Forensic analysis of social networking applications on mobile devices. *Digit Investig.* 2012;9(Suppl):S24-S33. doi:10.1016/j.diin.2012.05.007
 41. Ayers R, Brothers S, Jansen W. Guidelines on Mobile Device Forensics. NIST Special Publication 800-101 Revision 1. Gaithersburg, MD: National Institute of Standards and Technology. 2014. doi:10.6028/NIST.SP.800-101r1
 42. Alhassan JK, Oguntoye RT, Misra S, Adewumi A, Maskeliūnas R, Damaševičius R. Comparative evaluation of mobile forensic tools. *Proceedings of the International Conference on Information Technology & Systems (ICITS 2018).* Cham: Springer. 2018;105-114. doi:10.1007/978-3-319-73450-7_11
 43. Kong J. Data Extraction on MTK-based Android Mobile Phone Forensics. *Journal of Digital Forensics, Security and Law.* 2015;10(3):31-42. doi:10.15394/jdfsl.2015.1209

How to Cite: Sharma A, Yadav B, Sharma B. Evaluation of Commercial Mobile Forensic Tools for Android Evidence Extraction in Digital Forensic Investigations. *Int Res J Multidiscip Scope.* 2026;7(4):1696-1709. DOI: 10.47857/irjms.2026.v07i04.013962