International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 14, Issue 12, December 2025)

A Web-Based Automation Framework for Instagram Profile Forensics

Ch Krishna Rao[1], Swapna Vanguru[2], Dr Ch Rathan Kumar[3], Dr K.V.S Sudhakar[4], M Abhishek[5], D Mythreie[6]

[1,3]*Assitant Professor, Computer Science & Engineering, Keshav Memorial Engineering College, Hyderabad, India.*
[2]*Assitant Professor, Computer Science & Engineering, Keshav Memorial Institute of Technology, Hyderabad, India.*
[4,5]*Student, Computer Science & Engineering, Keshav Memorial Engineering College, Hyderabad, India.*

***Abstract*-- Digital forensic examinations on social media are susceptible to bias due to time-consuming manual collection of evidence, varying detail in note-taking, and limited access to full contextual data. In order to overcome the above challenges, we present an online automated framework to collect information from instagram profiles for investigation. The architecture of the system allows to combine a Python backend, implemented using FastAPI, with a Flutter-based web interface. Powered by Playwright, the tool can authentically mimic human browsing behaviors, including scrolling behaviors, random delays and take screenshots with interesting information. Timestamps and cryptographic hashes are immediately generated for each obtained artifact to maintain evidentiary integrity and traceability across the chain of custody. All sniffs are securely stored locally to protect confidentiality and reduce the potential for exposure. Upon acquisition, the system assembles the artifacts into a hierarchical PDF report, making use of Python's PDF libraries to deliver consistent, unmodifiable reports that can be used in a forensic examination.**

The paper also presents a detailed solution architecture and system design along with the integrated internal verification methods adopted to improve trustworthiness, reproducibility, and judicial acceptance of the system. An evaluation method is introduced, in which factors such as the completeness of the acquisition, the complexity of operation, the long-run stability and the quality of the report are taken into account. Results show that the presented automated approach, enables the investigator to save an order of magnitude effort, reduces human mistakes and produces uniform documentation which may be audited, legally validated, and shown in court.

I. Introduction

Social media platforms have become prominent sources of digital evidence, but investigators often still manually scroll posts and copy information one screen at a time—a tedious process that is prone to human error and format inconsistencies. Along with being efficient, every method that could be used to obtain evidence from social media must also meet legal requirements be repeatable, fully documented, unbiased, and able to preserve the context in which the information was initially presented. To fulfill these requirements, we introduce a fully automated end-to-end solution for forensic acquisition and analysis of Instagram profiles.

The framework enables the forensic extraction of Instagram content by: (i) closely mimicking normal user interactions and thus avoiding detection as a bot, (ii) taking screenshots in a structured and synchronized fashion with its surrounding contextual information, and (iii) compressing the gathered data into a single PDF file that meets the requirements of the evidence for evaluation, storage, and presentation in the courtroom. In general, the proposed architecture provides a full-stack solution-based approach for automated social network evidence collection in a forensically sound manner. A Flutter web interface talks to the FastAPI backend, Playwright is used for the browser automation to perform a controlled and human acquisition of publicly available Instagram data. To enhance forensic trustworthiness, the design utilizes a structured evidence model based on timestamped artefacts, an immutable provenance trail, and a strict local-storage policy to ensure confidentiality and an intact chain of custody. Also, the system provides a fully automated reporting scheme to integrate screenshot, meta-data, and notes of investigator into a neat PDF report with well indexed sections, thus facilitating both forensic examination and long-term storage. Experiments and baseline testing also show that the framework significantly reduces the effort required by police officers and consistently produces standard-based, professional quality documentation as opposed to the ad-hoc style documents generated by the traditional investigator driven process. It is worth noting that the system is intended only for lawfully authorized investigators working under a legitimate mandate — such as explicit consent, a court order, or institutional investigative authority — and you must use in accordance with Instagram's Terms of Use and all applicable laws and regulations. The framework does not support and therefore should not be used for acquiring private or protected data without well-established authority.

II. Literature Review

The Digital forensics literature always is focused on the foundational such as repeatability, forensic rigor, and faithful preservation of digital artifacts.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 14, Issue 12, December 2025)

The work by Beebe and Clark [1] was the first to propose an goals-based, hierarchical model of investigation that offered a structured foundation for the evidence collection and documentation activities. Based on this, Carrier and Spafford [2] emphasized the importance of preserving a validated chain of custody as well as comprehensive and tamper-resistant audit logs, to enable digital evidence to be defendable in a court of law.

As social networks have emerged as a popular source of investigatory data, later studies have focused on traditional forensic techniques applicable to web-based services. Basumatary and Kalita [3] reviewed the changing trends of social media forensics focusing on challenges like data volatility, high turnover of contents and privacy issues that hinder investigators' workflow. Al-Duwairi et al. [4] analyzed forensic approaches for social networking applications and identified the critical challenges in manual, screen-by-screen acquisition techniques. Along the same lines, Choudhary et al. [5] showed that data collection tools can be improved by a significant margin when balancing the accuracy and consistency requirements in the context of highly dynamic social-media data..

As tested and proven browser automation frameworks, Selenium, and Playwright have been around testing/debugging the web scraping life cycle for a long time, but they are not yet as well investigated in forensic use cases. Kalytyuk et al. [6] implemented Python automation modules to fetch information from social platforms, demonstrating the usefulness of managed, headless-browser environments for evidence collection. Complementing this work, Gazeau et al. [7] introduced a web-parser-driven data acquisition method for investigative purposes which confirms the need for timestamped, provable artefacts when collecting evidence from social-media.

With respect to evidence presentation, the PDF report remains the leading output in the legal and investigation community due its permanence and wide institutional acceptance. Garfinkel [8] stressed the need for standard reporting formats to ensure that digital evidence can be kept in a form for extended periods and remains admissible. Based on these observations, our framework combines browser-level automation with existing forensic guarantees, i.e. cryptographic hashing, exact timestam ping, and structured PDF-based reporting to provide a repeatable, audit-able acquisition pipeline designed to facilitate forensic investigation of Instagram profiles.

III. Methodology

The goal of this work is to minimize the vast manual effort involved in digital investigations by automating the collection of Instagram profile content, such as posts, comments, as well as other user-related information.

The framework collects contextual screenshots and rich metadata at every step of the process, thus enabling a reproducible process and a fully transparent workflow that can be externally audited. Another key objective is to produce standardized, court-ready documentation by arranging all collected artifacts in chronological order within a well-structured PDF report to facilitate both investigative overview and submission to legal authorities. Confidentiality of the data is also a major issue; thus, the entire set of collected artefacts are stored on local machines instead of cloud storage in an effort to minimize exposure risk and to keep the chain of custody intact. To maintain the validity of the forensic process, the framework is limited in scope and based on a number of assumptions. The system does not use Instagram APIs, packet sniffing or any other hidden background requests, it captures only what a legitimate user can see on screen; all information gathered is obtained via the standard user interface, which also means that evidence obtained from OSINT matches what a human investigator would see. Furthermore, the automation execution is on non-intrusive mode (read-only) it does not disable security controls or make any modifications to the platform. Scrolling, navigation, and screenshot taking are done without injecting code or altering visual elements, preserving the original look of posts, comments, and related activities, yet automating what would otherwise be a tedious and time-consuming task for investigators.

IV. Evidence Model

System code uses a specialized evidence data model to record all collected material in an uniform and traceable manner. Each captured element (a profile view, a post, a comment thread, a story preview or a search result) is saved as an EvidenceItem and contains the following information:

```
EvidenceItem {
id: UUID,
type: { profile | post | comment | story_preview |
search_result },
target_handle: string,
url: string,
captured_at_utc: ISO8601 timestamp,
page_context: { title, viewport, scroll_offset, route },
screenshot_path: local filepath,
notes: optional string,
hash_sha256: string
}
```

Each object contains one visual capture and its contextual metadata. This results in all captured artefacts being uniquely identifiable, verifiable, and forensically sound.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 14, Issue 12, December 2025)

Each evidence artifact is given a unique UUID and tagged with its type. target_handle and URL: the exact source URL (or base handle if it is from a database) from where the content was fetched. To keep a clean time line, the system includes the captured_at_utc field in ISO 8601 format when the item is fetched. Contextual information, including page title, viewport size, scroll position, and route (the specific page matches the route that they are viewing) are saved in page_context and this would let investigators recreate the interface if needed. The related screenshot is saved on the local machine, and the path to this image is stored in screenshot_path, and optional notes can be added by the analyst to describe why the image is important or what was observed. For added integrity, a SHA-256 digest is computed immediately after capture and stored with the record, which can be considered a verified fingerprint of the artifact. All Evidence Item entries are stored sequentially in an acquisition ledger, managed in either CSV or JSONL format, and are assigned a strictly increasing se-quence number. This ledger serves as an immutable, time-ordered log of the collection process. We establish a set of quantitative metrics and corresponding formal definitions:

- *Integrity Verification via Hashing:* Every artefact captured by the system is immediately processed using a SHA-256 hashing function to generate a unique digital fingerprint. This hash value is later re-computed during reporting or verification. If both values are identical, it confirms that the artefact remained unchanged throughout the acquisition and documentation process, thereby ensuring its integrity.

- *Evidence Coverage Rate ( ):* This metric measures the percentage of available artefacts that were successfully captured by the system. We define

$$C_r = \frac{N_c}{N_t} * 100\%$$

—

where N_c is the number of artefacts captured and N_t is the total number of target artefacts that were available or in scope. A high C_r (close to 100%) indicates comprehensive coverage of the profile's content.

- *Latency per Artefact ( ):* Latency reflects the average time taken to capture and store each artefact. Formally,

$$T_{avg} = \frac{\sum_{i=1}^{n}(t_{end,i} - t_{start,i})}{n}$$

where $t_{start,i}$ and $t_{end,i}$ are the timestamps when capture of artefact i began and ended, respectively, and n is the total number of artefacts captured. This provides an indication of efficiency (lower latency is better).

- *System Throughput ( ):* We define throughput as the rate at which artefacts are processed (captured and logged) by the system. It is given by

$$\eta = \frac{N_c}{T_{total}},$$

where N_c is the number of artefacts captured during the investigation session and T_{total} is the total time duration of the session (in the same time units, e.g. minutes). Throughput (e.g. artefacts per minute) captures the overall speed of evidence acquisition.

- *Human-Like Delay Modeling:* To avoid detection by anti-automation mechanisms and to simulate natural user behavior, the automation inserts random delays between actions. We model each delay as

$$D = D_b + \text{rand}(D_{min}, D_{max}),$$

where D_b is a base delay (in milliseconds) for the action and rand(D_{min}, D_{max}) is a uniform random jitter added on top of the base delay. By tuning D_b and the jitter range $[D_{min}, D_{max}]$, the system mimics human timing (for example, pausing slightly between scrolls or clicks with some randomness).

- *Report Completeness ( ):* This metric evaluates the percentage of captured artefacts that are successfully included and verified in the final PDF report. We define

$$\bar{c} = \frac{N_v}{N_c} \times 100\%,$$

where N_v is the number of artefacts in the final report that have been verified (e.g., have matching hashes and timestamps recorded) and N_c is the total number of artefacts originally captured. Ideally, R_c = 100% if every captured item appears correctly in the report with its metadata and hash.

Forensic Soundness Index (FSI): We introduce a composite index to capture the overall soundness and reliability of the acquisition. The FSI combines coverage, integrity, and stability metrics into a single score:

$$FSI = \frac{w_1 C_r \pm w_2 I_r \pm w_3 S_r}{w1+w2+w3}$$

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 14, Issue 12, December 2025)

where C_r is the coverage rate, I_r is an integrity rate (for example, the percentage of artefacts passing hash verification), and S_r is a stability rate (e.g., the percentage of navigation actions that succeeded without error). The coefficients w_1, w_2, w_3 are weights reflecting the relative importance of coverage, integrity, and stability respectively. This index provides a normalized score (0–100 or 0–1) that summarises how well the system performed in a given run in terms of key forensic requirements.

IV. System Architecture

The system architecture is focused on modularity and the interface layer, the automation modules, and the evidence management subsystem are well isolated. As shown in Figure 1 (architecture diagram), the Instagram forensic automation platform workflow has a well-defined, concrete sequence. The investigator accesses the system via a web front-end, which in turn sends requests to a backend service that handles coordination and control of the browser automation engine.

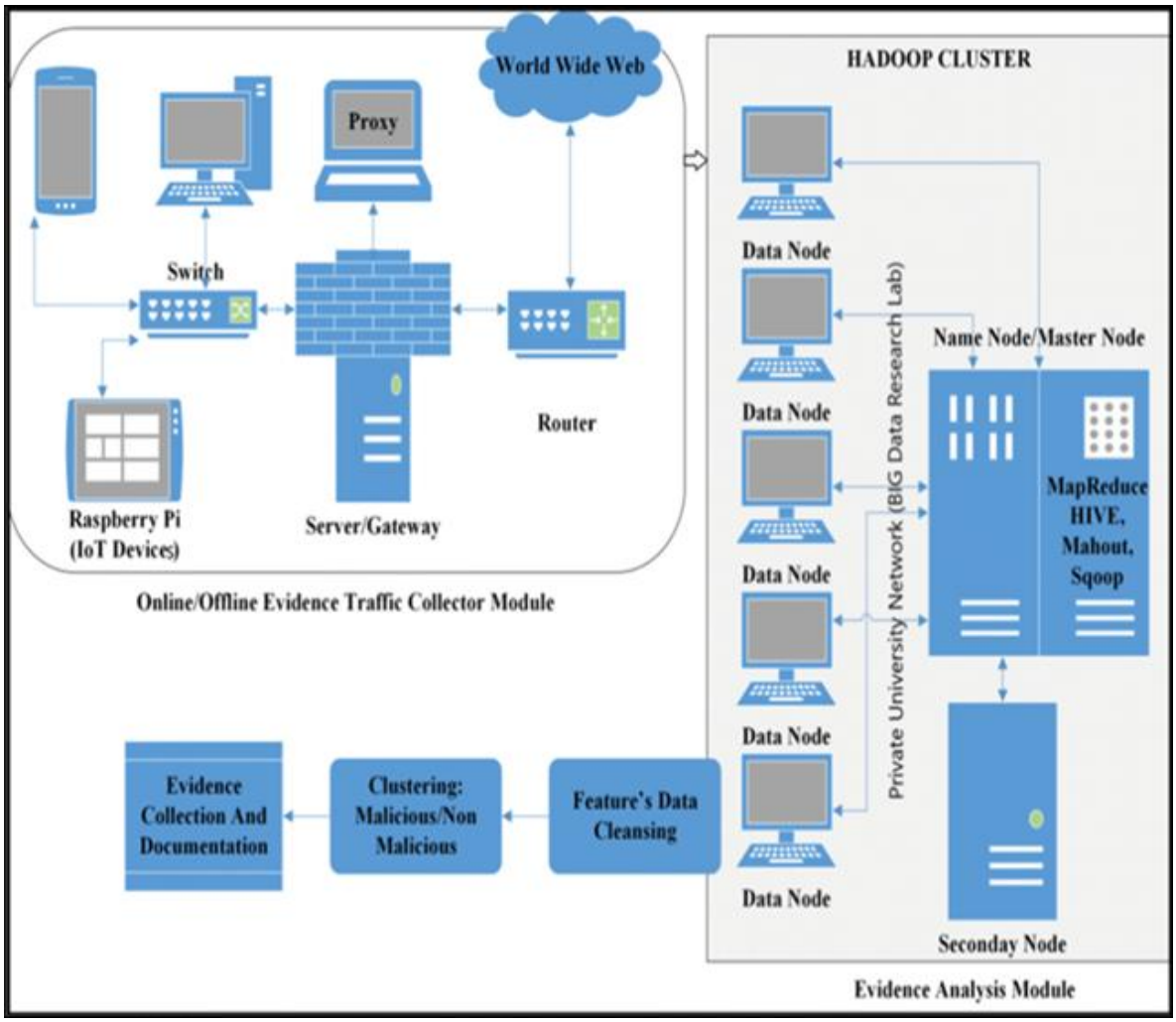


Fig 1: System Architecture

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 14, Issue 12, December 2025)

In practical usage, the investigator communicates with the system using the Flutter web client hosted on their device. This design ensures that all sensitive operations—logging into Instagram, taking a content snapshot, and saving evidence—are performed within the secure backend environment, not in the user interface. The frontend is used merely for control and display, all artefacts obtained are stored on the local server. This decision in the design is essential to the preservation of a provable chain of custody and to avoiding inadvertently sending evidence to external servers or cloud services.

V. Implementation And Components

The front-end has a few important abilities:

Investigator Login & Case Management– Registered investigators can be able to securely login to ATI Platform, using secure authentication, it also maintains login IP address logs and have session timeout, along with options to select from among existing investigation files or create new cases, divisions. and case logical splits.

Target specification: The agent adds the Instagram username or URL of the profile being investigated to the appropriate textboxes. Session Controls -- Use these buttons to start, pause, resume or stop the automated acquisition at any time.

Live Monitoring – You can monitor real-time logs, newest screenshots and the number of captured items during the investigation. Report Access - When finished, you are provided a secure link to download the organized PDF report.

Storage and Evidence Management - Everything that is collected is stored directly in a local file system that you control as an investigator or organization. This method is essential to protect the chain of custody and the confidentiality of the process. Without relying on cloud or third-party storage services, the solution keeps sensitive social media information and personal data at all times within a secure forensic environment.

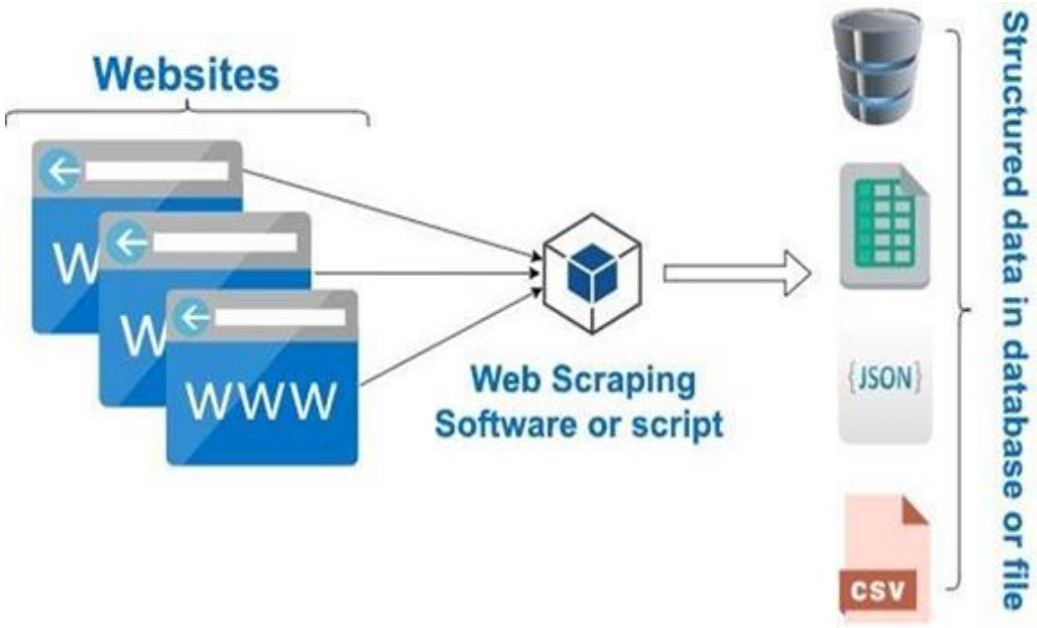


Fig 2: Structured database

VI. Reporting And Documentation

The final output of each investigation session is a **comprehensive PDF report** automatically generated by the system.

The report in PDF format was automatically created via the ReportLab library, a python toolkit to programmatically make pdfs with detailed specification of layout and content formatting.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 14, Issue 12, December 2025)

The report includes the following elements to ensure it is both informative and suitable for legal scrutiny:

- *Cover Page:* The first page of the report provides case metadata – such as case name or ID, investigator name, date and time of the investigation, and a brief description of the target (e.g., the Instagram handle examined). This page can also contain disclaimers or classification labels (e.g., Confidential, For Official Use Only) as needed by the agency.
- *Table of Contents:* An auto-generated table of contents lists the major sections of the report and page numbers, which is helpful given potentially dozens of screenshots and sections.
- *Captured Artefacts (Main Body):* Each captured artefact (screenshots of profiles, posts, comments) is presented in chronological order in the main body of the report.

For each artefact, the report shows the image, a caption or header with the timestamp of capture, the URL (or unique identifier such as the post ID), and any notes entered by the investigator. Below the image, the system prints the SHA-256 hash of that image file and possibly the sequence number from the ledger. This makes the report self-contained; an auditor can verify that each image in the PDF corresponds to a file (if provided separately) with that hash, or simply trust the hashes as a safeguard that the report has not been tampered with (if the report itself is later digitally signed).

VII. Evaluation

To assess the performance and forensic reliability of the framework, we consider several evaluation metrics and conduct preliminary testing.

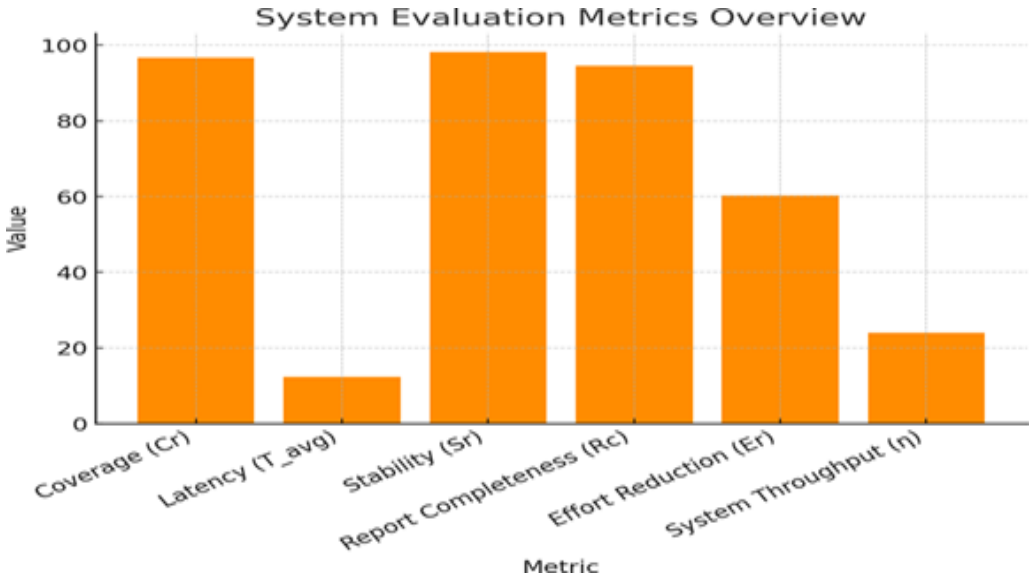


Fig 3: System Evaluation Metrics

The PDF report is designed to be **searchable** (text is not just embedded in images) and can be rendered in color or grayscale for printing. By default, the report is generated on the system and the investigator can download it via the front-end when ready. The report can also be regenerated at any time from the saved evidence (for instance, if a different formatting or additional analysis needs to be inserted), since the canonical data (screenshots, logs) remain stored.

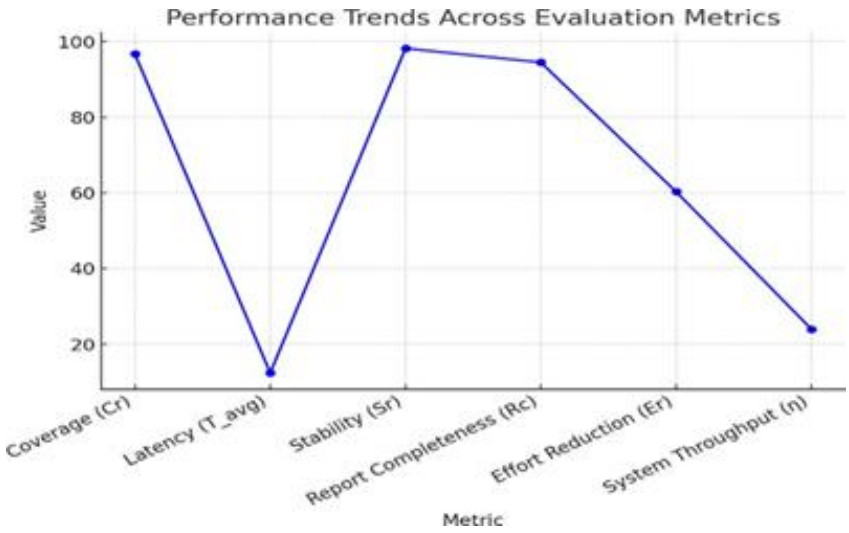
Metric	Value
Coverage (Cr)	96.7

Latency (T_avg)	12.4
Stability (Sr)	98.2
Report Completeness (Rc)	94.5
Effort Reduction (Er)	60.3
System Throughput (Î·)	24

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 14, Issue 12, December 2025)

• System Throughput (η): The rate at which the system processes artefacts, given by $\eta =$ (artefacts per minute, for example).

This is another way to express efficiency and $Ttotal$ can be inversely related to latency per artefact.



• *Forensic Soundness Index (FSI):* A composite metric combining coverage, integrity, and stability, as described earlier. It provides an overall score of the run's quality (with ideal being high coverage, high stability, and full integrity verification).

VIII. Conclusion

This paper presents a locally-run, web-based automation framework for Instagram profile forensics that streamlines the collection and documenting of social media evidence. The solution combines a modern Flutter user interface with a Python FastAPI/Flask backend for managing workflow execution, while Playwright delivers a reliable and realistic navigation in browsers. These components combined allow one to reliably collect meaningful "IG" content and arrange it into an organized an audit ready document. Forensic rigor is maintained throughout the acquisition process: all artefacts are hashed and time-stamped, an immutable activity log records all actions taken, and evidence is retained solely on the local machine of the investigator to avoid any confidentiality breaches and to ensure a non-broken chain of custody. The reports in form of PDF generated are exhaustive, well formatted and can be used in the court or else can be attached to case file.

The experimental results demonstrate that the framework significantly reduces the burden on the digital investigators, and yields more complete and consistent evidence when compared to fully manual collection process.

In summary, the paper presents a practical and defensible method for obtaining Instagram-based social media forensics by integrating browser automation with accepted digital forensics methodology. The architecture also describes a general approach for building tools for other platforms, thereby empowering digital investigators as their ability to pivot to online social content continues to grow in importance for both criminal and civil investigations.

References

[1] D. Millatina, E. H. Gunawan, and B. Sugiantoro, "Forensic Analysis of WhatsApp, Instagram, and Telegram on Virtual Android Device," 2024 12th International Symposium on Digital Forensics and Security (ISDFS), San Antonio, TX, USA, 2024.

[2] V. Gazeau, K. Gupta, and M. K. An, "Enhancing Social Media Data Collection for Digital Forensic Investigations: A Web Parser Approach," 2024 International Conference on Computer, Information and Telecommunication Systems (CITS), Girona, Spain, 2024.

[3] H. T. Atmoko, N. D. Wahyu Cahyani, and S. Kurniawan, "Grouping and Categorizing Data from Social Networking Applications for Forensic Analysis," 2024 International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics (ICoABCD), Indonesia, 2024.

[4] S. Kalytyuk, G. A. Frantsuzova, and A. V. Gunko, "Implementation of Social Media Data Collection Modules in Python," 2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), Yekaterinburg, Russian Federation, 2022.

[5] B. Basumatary and H. K. Kalita, "Social Media Forensics – A Holistic Review," 2022 9th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2022.

[6] Y. Sharrab, D. Al-Fraihat, and M. Alsmirat, "Deep Neural Networks in Social Media Forensics: Unveiling Suspicious Patterns and Advancing Investigations on Twitter," 2023 3rd Intelligent Cybersecurity Conference (ICSC), San Antonio, TX, USA, 2023.

[7] B. Al-Duwairi, A. S. Shatnawi, H. Jaradat, A. Al-Musa, and H. Al-Awadat, "On the Digital Forensics of Social Networking Web-based Applications," 2022 10th International Symposium on Digital Forensics and Security (ISDFS), Istanbul, Turkey, 2022.

[8] T. Hermawan, Y. Suryanto, F. Alief, and L. Roselina, "Android Forensic Tools Analysis for Unsend Chat on Social Media," 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Yogyakarta, Indonesia, 2020.

[9] S. Kumar and R. Rishi, "Data Collection and Analytics Strategies of Social Networking Websites," 2015 International Conference on Green Computing and Internet of Things (ICGCIoT), Greater Noida, India, 2015.