INTERNATIONAL JOURNAL of SMART GRID
L. Rzayeva et al., Vol.9, No.3, September, 2025

Secure Chip-Off Method with Acoustic-based Fault Diagnostics for IoT and Smart Grid Data Recovery

Leila Rzayeva *, Abilkair Imanberdi **, Aigerim Alibek ***, Ali Myrzatay****, Yerassyl Yermekov*****‡, Korhan Kayisli******, Gerald Feldman*******

* Research and Innovation Center "CyberTech", Astana IT University, Astana 010000, Kazakhstan;
** Dean's Office, Astana IT University, Astana 010000, Kazakhstan;
*** Research and Innovation Center "CyberTech", Astana IT University, Astana 010000, Kazakhstan
**** Research and Innovation Center "CyberTech", Astana IT University, Astana 010000, Kazakhstan;
*****Department of Information security, Faculty of Information Technologies, L.N. Gumilyov Eurasian National University, Astana 010000, Kazakhstan;
****** Gazi University, Eng. Fac. Electrical-Electronic Eng. Ankara, Turkiye; korhankayisli@gmail.com
******* Birmingham City University (Faculty of Computing, College of Computing);
(l.rzayeva@astanait.edu.kz, a.imanberdiyev@astanait.edu.kz, a.zhenisbekkyzy@astanait.edu.kz, mirzataitegiali@gmail.com, yerassylyermekov@gmail.com, korhankayisli@gmail.com, Gerald.Feldman@bcu.ac.uk)

‡ Corresponding Author y.yermekov@gmail.com, 010000, Tel.: +7 706 639 2930.
Received: 18.06.2025 Accepted:24.07.2025

Abstract- This article explores modern methods for extracting information from faulty mobile devices, hard disk drives (HDDs), and solid-state drives (SSDs) while considering the physical integrity of data storage components. In the digital era, recovering data from damaged devices is crucial for forensic investigations, corporate security, and information protection. The study examines existing data extraction techniques for mobile devices, including both software-based and hardware-based approaches such as JTAG, SPI, UFI Box, and the "Chip-off" method. It highlights the importance of low-level data access; as logical extraction methods often fail to recover deleted or hidden files. For HDDs, the paper classifies possible failures into logical and physical damage categories. It discusses data recovery mechanisms, ranging from diagnosing disk health and analyzing SMART attributes to utilizing specialized recovery tools and hardware techniques, such as replacing the magnetic head assembly (MHA) and reconstructing the file system. Additionally, the work incorporates an Environmental Sound Recognition (ESR) module to enable the automated detection of mechanical failures based on acoustic signatures. As the adoption of IoT devices with onboard storage accelerates, ensuring secure, reliable, and forensic-ready data recovery methods becomes increasingly important. The proposed chip-off method with acoustic diagnostics supports critical security and privacy needs in IoT ecosystems by enabling recovery and analysis of compromised or tampered edge devices. The research contributes to the advancement of forensic analysis and data recovery techniques, offering valuable insights for law enforcement agencies, private investigators, and cybersecurity professionals. This methodology not only enhances forensic capabilities but also supports data recovery within secure smart grid environments and IoT-based infrastructures, where device tampering and data breaches are critical concerns.

Keywords Smart Grid Security, IoT Data Integrity, AI Diagnostics, environmental sound recognition (ESR), acoustic diagnostics, physical damage, magnetic head assembly (HSA)

1. Introduction

In the realm of digital forensics and data recovery, hard disk drives (HDDs) remain a primary medium for storing information; however, they are vulnerable to both logical and physical failures, which can result in significant data loss for individuals and organizations. Beyond accidental damage, deliberate tampering or destruction of HDDs is increasingly observed in anti-forensic scenarios, where actors employ techniques such as data obfuscation, artifact erasure, or even physical destruction to impede investigations. Traditional chip-off methods, where the storage controller is removed to access raw memory chips, have proven effective in retrieving low-level data when conventional logical extraction fails. However, these processes often rely on manual diagnostics, including auditory inspection ("check the sound") to identify mechanical faults, which can be time-consuming and subjective [1-3].

Concurrently, advances in environmental sound recognition (ESR) have demonstrated that machine learning algorithms, particularly Support Vector Machines (SVMs), can accurately classify complex acoustic patterns from varied audio sources. ESR techniques, ranging from handcrafted feature extraction (e.g., MFCCs, chroma, spectral contrast) to deep learning–based models, have been successfully applied in domains such as urban monitoring, smart home automation, and surveillance. The robustness of these audio-based classifiers to background noise and their ability to distinguish subtle differences in sound events suggest a promising avenue for automating HDD fault diagnostics [12-15].

In addition, as the Internet of Things (IoT) ecosystem continues to grow, an increasing number of devices rely on local data storage using HDDs or SSDs. Ensuring the security and privacy of data on these storage components is critical, particularly when IoT devices become targets of physical tampering or cyber-attacks. Forensic-grade recovery and diagnostics, such as the proposed method, are vital for investigating breaches and maintaining data integrity in IoT environments.

Building on these parallel developments, this study proposes a novel Chip-Off Method for data extraction from HDDs incorporating acoustic-based fault diagnostics. By integrating ESR algorithms into the chip-off workflow, we aim to replace manual auditory assessment with an automated, data-driven classification of drive health. This approach not only enhances diagnostic objectivity and repeatability but also accelerates the identification of mechanical anomalies, such as head clicks, motor grinding, or spin failures, thereby streamlining the overall data recovery pipeline [4-6].

The scientific contributions of this study include the following:

- Development of a new chip-off method for data extraction from HDDs with physical or logical damage, tailored for forensic applications.
- Integration of an automated Environmental Sound Recognition (ESR) module into the diagnostic phase of the chip-off workflow.
- Creation of a structured, phase-based HDD acoustic dataset enabling fault detection across startup, idle, and load phases.
- Implementation of a reproducible methodology combining S.M.A.R.T. telemetry, acoustic analysis, and hardware intervention in a unified decision pipeline.

In smart grid and IoT ecosystems, the ability to resist cyber-physical attacks and preserve high-quality data integrity is essential. The proposed methodology contributes to this goal by offering a robust and secure means of recovering critical data from compromised storage media, thus reinforcing digital trust in next-generation power systems. While the primary motivation of this research lies in digital forensics and recovery from various failure modes, the proposed methods are also highly relevant to critical infrastructure systems such as smart grids. Edge devices in substations and smart meters often rely on embedded storage that may suffer damage due to power surges, tampering, or targeted cyber-attacks. Techniques explored in this paper, such as chip-off analysis and low-level diagnostics, can support integrity verification and rapid recovery in such scenarios.

2. Literature Review

Hard disk drive (HDD) data recovery is essential in digital forensics, cybersecurity, and information protection. Conventional extraction methods often fail on devices with physical damage or limited access, underscoring the need for advanced approaches such as the chip-off technique. In chip-off, the memory chip is physically removed from the HDD and read directly, bypassing damaged components. However, this procedure requires exceptional precision, as execution errors can result in irreversible data loss [7-8].

Recent studies on chip-off have focused on refining extraction protocols to enhance efficiency and minimize media damage. Literature describes techniques ranging from delicate solder-based chip removal to using specialized readers that can access data without full desoldering. Despite these advances, significant challenges remain, particularly with severely damaged drives where the nature of the fault is unclear, complicating the choice of recovery strategy [9-11].

Concurrently, HDD fault diagnosis research has emphasized acoustic methods that detect mechanical defects by analyzing the sounds produced during operation. Deviations in acoustic signatures can indicate issues such as read/write head malfunctions or spindle motor failures. These non-invasive approaches can complement existing recovery procedures by providing preliminary insights into disk health [12-13].

Environmental Sound Recognition (ESR) makes a substantial contribution to acoustic analysis. ESR utilizes machine learning algorithms, including Support Vector Machines (SVM), to classify audio signals. Using libraries like Librosa, researchers extract features including Mel-Frequency Cepstral Coefficients (MFCCs), chroma vectors, and mel-spectrograms. These methods have proven effective in categorizing diverse environmental sounds and are readily adaptable for HDD diagnostics through acoustic profiling [14].

Integrating acoustic fault diagnosis with chip-off methods opens new avenues for more effective data recovery. Preliminary acoustic classification of the fault type allows tailoring the chip-off process to the specific condition of the damaged HDD, potentially increasing success rates and reducing further harm. Nonetheless, this fusion remains underexplored in literature, highlighting gaps and the need for additional development [15-17, 23].

Early approaches to Environmental Sound Recognition (ESR) commonly depended on the manual derivation of audio descriptors such as Mel-Frequency Cepstral Coefficients (MFCCs) and various spectral characteristics. These features were subsequently input into classical machine learning classifiers, including Support Vector

Machines (SVMs) and k-Nearest Neighbors (k-NN). The effectiveness of such systems was highly contingent on the discriminative power and reliability of the extracted features, which are susceptible to distortion due to ambient noise, recording conditions, and the inherent variability of sound sources. Consequently, traditional methods often lacked the flexibility and robustness needed for consistent performance across heterogeneous real-world datasets [16-18].

In recent years, ESR has undergone a significant transformation due to the adoption of deep learning techniques, which allow for end-to-end learning of features and classification tasks directly from raw audio waveforms or spectrograms. Architectures such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks have had a particularly significant impact. CNNs excel at identifying spatial and temporal patterns in spectrograms, while LSTMs are well-suited for capturing temporal dependencies and sequence dynamics. This evolution toward data-driven modeling has enabled the construction of scalable and highly accurate ESR systems, as demonstrated by successful applications in domains such as urban acoustic scene analysis and wildlife sound monitoring [19-20]. Moreover, the use of modern enhancement strategies such as transfer learning and data augmentation has further improved ESR performance. Transfer learning leverages models pre-trained on large-scale datasets, which can be fine-tuned for specific ESR tasks with relatively small labeled datasets, reducing training costs. Simultaneously, augmentation techniques—like pitch modulation, time warping, and synthetic noise injection—have proven effective in increasing data diversity, thereby enhancing generalization and model resilience.

Despite these technological advances, several obstacles persist. Environmental sounds often exhibit significant variability and can co-occur with other audio events, complicating the classification task. Background noise, reverberation, and device constraints also present practical limitations, particularly in real-time or embedded ESR applications. Overcoming these issues requires continued refinement in audio representation methods, neural architectures, and learning frameworks [21]. As a result, an enhanced chip-off technique for recovering data from damaged hard disk drives (HDDs) is proposed. As part of the diagnostic process, machine learning-based methods are integrated to perform acoustic analysis of HDD failure sounds. These auxiliary techniques enable automated classification of drive malfunctions based on audio recordings, providing additional insights before physical intervention. By employing Environmental Sound Recognition (ESR) approaches, specifically Sound Recognition (ESR) approaches, Support Vector Machines (SVMs) trained on features such as Mel-Frequency Cepstral Coefficients (MFCCs), chroma vectors, and Mel spectrograms, introduces a non-invasive pre-assessment step. While not central to the recovery pipeline, this integration improves the precision of fault identification and helps optimize the chip-off workflow.

In practical deployments, field recordings with background noise or limited acoustic quality can significantly degrade the ESR system's performance. Distorted or masked audio features may lead to confusion between fault types or reduce the confidence level of classification. To address this, we are currently exploring strategies such as spectral enhancement, denoising filters and data augmentation techniques—including time stretching, pitch shifting, noise injection. These approaches aim to improve the ESR model's robustness to environmental variability and increase its diagnostic reliability in real-world forensic scenarios.

3. Methodology

3.1 Information Extraction Methods and Sound-Based Fault Classification

The process of recovering data from a malfunctioning hard disk drive (HDD) involves multiple diagnostic and technical stages. In this study, a combined recovery strategy was applied, integrating physical chip-off procedures with a supplementary acoustic analysis step.

The workflow begins with device inspection, where visual and electronic indicators help determine the extent of physical damage. If the system recognizes the drive upon connection, further actions are guided by a predefined recovery algorithm. Modern HDDs often support S.M.A.R.T. (Self-Monitoring, Analysis and Reporting Technology), which provides preliminary diagnostic insights. In cases where system-level access is possible, the disk is connected and analyzed through software tools. Otherwise, the hardware-level inspection proceeds, including the removal of protective elements and the potential replacement or repositioning of internal components, such as the head assembly. Based on the created data recovery algorithm, the process has now reached step 3 (Fig. 1).

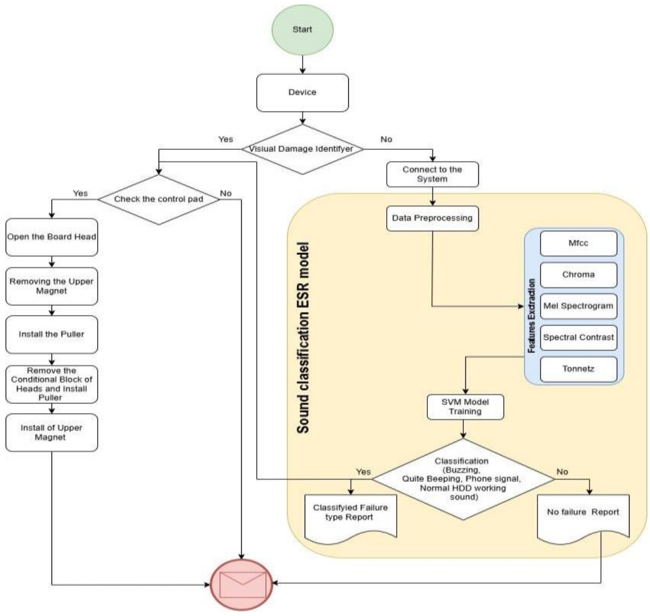


Fig. 1. The proposed HDD data recovery approach.

To improve early fault detection and reduce unnecessary physical intervention, an Environmental Sound Recognition (ESR) module was integrated into the initial diagnostic phase. The ESR model operates by capturing and analyzing acoustic signals generated by the HDD during startup or operation. Key audio featuressuch as Mel-Frequency Cepstral Coefficients (MFCCs), chroma vectors, and Mel spectrograms, are extracted using the Librosa library and passed into a Support Vector Machine (SVM) classifier trained to detect known failure signatures. This acoustic analysis step results in one of two outcomes:

- If no abnormal patterns are detected, the device is considered mechanically stable, and physical disassembly may be avoided.Integration of an automated Environmental Sound Recognition (ESR) module into the diagnostic phase of the chip-off workflow.
- If failure-specific sound signatures are identified, a failure type report is generated to guide the subsequent chip-off process with improved precision.

The ESR framework uses methodologies that are part of the larger research work undertaken as part of this project, and hence extends the application of those methodologies. The earlier work focuses on the feature extraction process, model architecture, and performance evaluation of SVM classifiers applied to structured environmental sound datasets. In this current context, those principles are adapted to the domain of HDD fault acoustics.

3.2 Materials

1) Acoustic Fault Classification via ESR

To develop and adapt the ESR classifier for HDD diagnostics, we followed a two-stage training approach. In a previous study, we trained a baseline Support Vector Machine (SVM) model using the ESC-10 dataset, a curated 10-class subset of the widely used ESC-50 dataset of environmental sounds dataset retrieved from [22]. This dataset comprises 400 labeled audio recordings from various real-world sound categories and was used to validate the feasibility of feature extraction pipelines and the general structure of the classifier, utilizing descriptors such as MFCCs, chroma vectors, and Mel spectrograms, as shown in Figure 2.

In the current work, a specialized audio dataset was created by recording each HDD five times under controlled conditions. Each recording captured the entire operational cycle of the disk, from power-on to power-off. This enabled the transfer and adaptation of the learned model to a custom dataset of 36 HDDs, which included 11 non-functional units. To improve granularity in fault classification, each recording was segmented into three distinct time windows:

- Startup phase: initial spin-up and head positioning;
- Idle phase: passive rotation or standby state; and
- Load phase: simulated I/O operations or stress tests to engage read/write mechanisms.

This structure allowed the model to distinguish between acoustic anomalies that manifest only during specific phases of operation. For example, startup-phase clicking may indicate mechanical obstruction, while noise underload may point to degraded seek performance or platter issues. The model was retrained and fine-tuned using this 36-drive dataset to adapt to the unique acoustic signatures of HDDs. To further enhance the training process, data augmentation techniques were employed. These included:

- Time Stretching: Modifying the speed of the audio without affecting its pitch to create variations.
- Pitch Shifting: Changing the pitch of the audio to simulate different sound frequencies.
- Adding Background Noise: Incorporating various levels of background noise to make the model robust to real-world conditions.

These augmentation methods help prevent overfitting and improve the model’s ability to generalize to new, unseen data. The dataset was split into training and testing sets using an 80-20 split ratio, meaning that 86 samples were used for training the model. In contrast, the remaining 22 samples were reserved for testing its performance. This split was performed in a stratified manner to ensure that each class was proportionally represented in both the training and testing sets.

A Support Vector Machine (SVM) was used for the new classification task. SVMs are effective for high-dimensional spaces and suitable for small to medium-sized datasets. Feature extraction is performed using the ‘librosa‘ library, which includes:

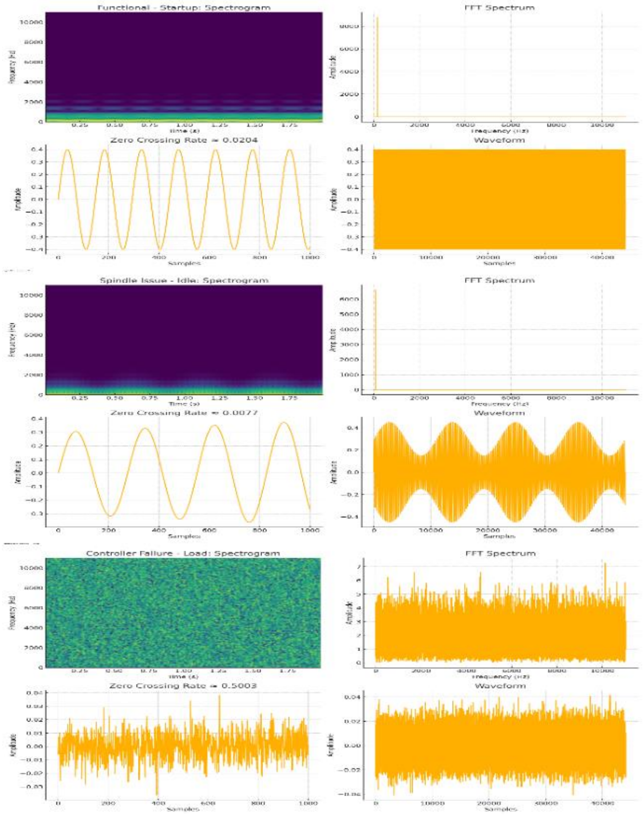
- MFCC: Mel-Frequency Cepstral Coefficients capturing the short-term power spectrum of sound,
- Chroma: Representing the twelve different pitch classes,
- Mel Spectrogram: Describing the power spectrum of the audio signal,
- Spectral Contrast: Highlighting the difference between peaks and valleys in the spectrum,
- Tonal Centroid Features (Tonnetz): Mapping tones to a six-dimensional space.
- FFT Spectrum (Fast Fourier Transform): a spectrum showing the energy (amplitude) in each frequency range.

The extracted features are then used to train the SVM classifier with a linear kernel.

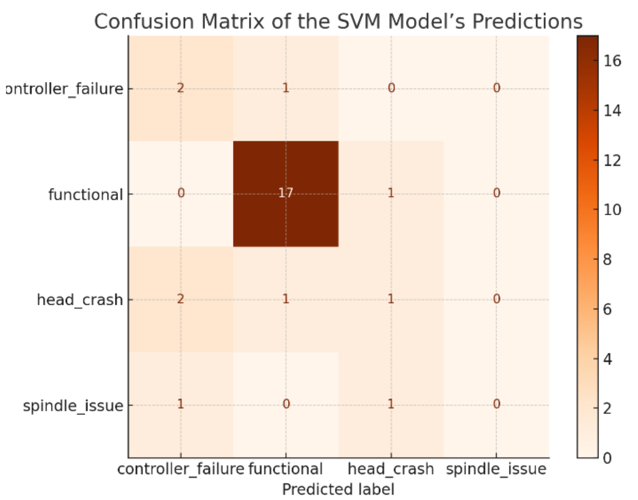
For feature extraction, librosa library has been used with written steps:

1. Load the audio file using librosa.load function.
2. Compute the Short-Time Fourier Transform (STFT) of the audio signal.
3. Extract MFCCs, Chroma, Mel Spectrogram, Spectral Contrast, FFT, and Tonnetz features.
4. Concatenate these features into a single feature vector.

To illustrate the diversity of acoustic signatures across HDD conditions, Figure 2a and Figure 2b present time–frequency and amplitude-domain visualizations of three representative audio samples. Each row corresponds to a specific disk state and operational phase—functional during startup, spindle degradation during idle, and controller failure during load.



(a)



(b)

Fig. 2. Visual analysis and classification results for HDD acoustic diagnostics.

(a) Comparison of feature representations extracted from three HDD audio recordings in different operational states: functional (startup), spindle issue (idle), and controller failure (load). Each panel shows the spectrogram, spectral energy distribution, waveform, and zero-crossing patterns.

(b) Confusion matrix of the SVM model's predictions on acoustically augmented HDD recordings.

The spectrograms reveal distinct energy distributions: functional drives exhibit clean and periodic frequency components, while degraded spindles show fluctuating power bands, and controller failures result in minimal or noisy signals. FFT plots emphasize the degradation of harmonic structure, while zero-crossing rates and waveform patterns support phase-specific anomaly detection. These variations confirm the feasibility of using feature-based ESR for pre-classification of HDD failures.

Model Training: After extracting the features, the dataset is split into training and testing sets using the train_test_split function from scikit-learn. The SVM classifier with a linear kernel is then trained on the training set.

The steps for model training include:

➢ Splitting the data into training and testing sets with an 80-20 split.

➢ Initializing the SVM classifier with a linear kernel.

➢ Training the classifier on the training set using the fit method.

Evaluation: After training, the model's performance is evaluated on the testing set. The primary metric used for evaluation is accuracy, which is calculated as the ratio of correctly predicted instances to the total instances.

Additional metrics such as precision, recall, and F1-score can be computed to provide a more comprehensive assessment of the model's performance.

The performance of the ESR system is evaluated using the following metrics:

➢ Accuracy: The ratio of correctly predicted instances to the total instances.

➢ Precision: The ratio of correctly predicted positive observations to the total predicted positives.

➢ Recall: The ratio of correctly predicted positive observations to all observations in the actual class.

➢ F1-Score: The weighted average of Precision and Recall, providing a balance between the two.

These metrics provide a comprehensive assessment of the model's classification capabilities.

Figure 2b presents the confusion matrix of the Support Vector Machine (SVM) classifier evaluated on a modified dataset of HDD audio recordings with some noise, phase perturbations, and some label inconsistencies to approximate real-world acoustic conditions. The matrix reveals strong performance in identifying functional drives, while error-prone categories such as controller failure and spindle issue demonstrate overlap due to their low acoustic distinctiveness. These results highlight the challenge of fault classification in realistic settings and emphasize the importance of robust feature design and dataset variability for ESR-based diagnostics.

The similarity in acoustic signatures between controller failure and spindle issues is likely due to their overlapping frequency components in the mid-range spectrum (1–3 kHz), which are common to both vibration patterns and motor-related anomalies. These shared characteristics reduce the discriminative power of basic features and increase the likelihood of misclassification. Such ambiguity can affect the recovery process by leading to inaccurate diagnostic decisions, such as unnecessary head replacement or delayed identification of the actual fault. In future work, we plan to explore more advanced feature extraction methods, including harmonic envelope modeling and attention-based neural classifiers to improve fault separation and classification accuracy in acoustically similar cases.

3.3 Analyzing of physical damage

The physical damage was analyzed using the following steps. First, inspect the drive controller board for any deformed, missing, or burnt elements, and verify the integrity of the connectors. If serious damage or burnt elements are detected, it is recommended not to supply power to the affected disk in order to prevent the problem from worsening.

Fig. 3. Checking the status of the pads on the controller board.

Next, the controller board is carefully detached, allowing the condition of the contact pads that link the PCB to the magnetic head assembly (HSA) to be examined. As illustrated in Figure 3, these contact zones are critical for ensuring electrical continuity between subsystems.

Fig. 4. Contacts connecting the controller board to the magnetic head unit (HSA).

In cases where oxidation is visible (as depicted in Figure 4), it can be cleaned using a standard pencil eraser, provided the contact surface is flat. For recessed or non-uniform areas, such cleaning methods are not advised. Figure 5 highlights a white sealing compound inside the HDD chassis; if disrupted or cracked, it typically indicates mechanical stress or prior unauthorized disassembly. Additionally, dust or moisture-related contamination may occur on the platters or internal surfaces. In such cases, isopropyl alcohol should be used for cleaning, and microscopic tools are recommended for precise inspection and targeted cleaning.

Fig. 5. The flooded white area inside the HDD enclosure.

Suppose the control board needs to be replaced. In that case, specific onboard components, such as the microcontroller unit (MCU), EEPROM, NV-RAM, or NAND, may need to be transplanted to a donor board, depending on the architecture. In many instances, transferring only the EEPROM is sufficient for compatibility. Critical attention must be given to the part numbers and layout of the donor board: the PCB's identifier, as well as the MCU and VCM/SM controller specifications, must match the original. Mismatched components can lead to electrical failure, including burnout of the preamplifier circuit.

Before reassembly and power application, a multimeter should be used to verify that both the 5V and 12V lines are free from short circuits. Resistance measurements across the motor windings are also essential; abnormal readings may indicate a malfunctioning preamp. Once electrical integrity is confirmed, the device can be safely reconnected and brought to the next stage of the recovery process.

Fig. 6. The engraved number of the printed circuit board.

After applying the power supply (Fig. 6), the second stage of analysis is acoustic fault classification via the ESR module. After applying the report from the ESR module, if any damaged HDD is found, the damage should be confirmed, for example, by opening the disk. Once the damage is confirmed, the magnetic head unit should be replaced. However, replacing a damaged magnetic head unit requires strict adherence to cleanroom protocols and the use of specialized tools. The process involves a sequential series of mechanical operations aimed at carefully extracting the defective head assembly and installing a functional donor component without damaging the platters or the internal alignment. All steps are performed according to the algorithm (Fig. 1).

Fig.7. Removing the upper magnet.

As depicted in Figure 7, the internal structure of the hermetic block becomes accessible once the cover is removed. At this stage, the upper magnet must be carefully detached using specialized tools to allow further disassembly of the actuator system. The extractor must be fixed with the magnetic head assembly to ensure stability, after which the damaged head unit can be safely lifted from the platter surfaces. An identical procedure is carried out on the donor drive to obtain a compatible and fully functional magnetic head block. Once verified, the donor head assembly is carefully installed into the target drive that requires data recovery.

After proper positioning, the head unit is aligned to the designated parking zone adjacent to the spindle, completing the mechanical installation. In the next step, the upper magnet is carefully reinstalled to ensure proper alignment with the actuator assembly. Finally, the hermetic enclosure is sealed by closing the lid, thereby completing the replacement of the magnetic head unit. At this point, the drive is mechanically reassembled and ready to proceed to the subsequent stage of diagnostics or data extraction.

Smart grid infrastructures consist of distributed intelligent electronic devices (IEDs) which are vulnerable to both cyber and physical attacks. In the event of system compromise or suspected tampering, methods such as ESR and chip-off provide an opportunity to perform secure forensic extraction of system logs and firmware. For example, if an edge device fails in a critical substation, chip-off analysis allows engineers to recover last-state operational data and investigate anomalies, even in cases of malicious destruction. The ESR method supports safe data inspection from embedded NAND/NOR memory without altering the original data — a key requirement for legally admissible digital evidence in smart grid breach investigations [24-26].

4. Results and Discussion

In this work, a custom diagnostic script was used to analyze S.M.A.R.T. attributes, enabling the early detection of potential hardware issues and the assessment of their severity. For instance, a rise in the number of reallocated sectors may indicate developing surface degradation, while sudden changes in seek error rates can point to actuator instability. In addition to S.M.A.R.T. data, an Environmental Sound Recognition (ESR) module was integrated to support automatic identification of mechanical failures based on acoustic patterns. This module classifies faults such as head crashes, spindle malfunctions, or controller failures by analyzing audio recordings from the drive during various operational phases. The combination of S.M.A.R.T. telemetry and ESR-based audio classification provides a more comprehensive diagnostic view, enabling condition-aware decision-making before initiating invasive recovery procedures.

This script (fig. 8) is used to obtain and display disk status information using psutil and smartctl. This includes disk partitioning, usage, and S.M.A.R.T. attributes. get disk info(): Gathers disk details and calls helper functions for S.M.A.R.T. data. Script details:

- get disk info(): Collects information about the disk and calls functions to help with S.M.A.R.T. data.
- get smart status(device): Run smartctl -H to determine the health status.
- get smart attributes(device): Run smartctl -A to restore
- S.M.A.R.T. attributes.

```
def get_smart_status(device):
    try:
        result = subprocess.run(["smartctl", "-H", device], capture_output=True, text=True)
        if "PASSED" in result.stdout:
            return "Healthy"
        elif "FAILED" in result.stdout:
            return "Failed"
        else:
            return "Unknown"
    except FileNotFoundError:
        return "smartctl not installed"

def get_smart_attributes(device):
    try:
        result = subprocess.run(["smartctl", "-A", device], capture_output=True, text=True)
        return result.stdout if result.stdout else "No SMART attributes found"
    except FileNotFoundError:
        return "smartctl not installed"

def display_disk_info():
    disks = get_disk_info()
    print("\n┌──────────────────────────────────────────────┐")
    print("│                Disk Information              │")
    print("└──────────────────────────────────────────────┘\n")
    for disk in disks:
        print(f"Device: {disk['Device']}")
        print(f"Mountpoint: {disk['Mountpoint']}")
        print(f"File System: {disk['File System']}")
        print(f"Total Size: {disk['Total Size'] / (1024**3):.2f} GB")
        print(f"Used: {disk['Used'] / (1024**3):.2f} GB")
        print(f"Free: {disk['Free'] / (1024**3):.2f} GB")
        print(f"Usage: {disk['Usage Percentage']}%")
```

Fig. 8. Script for smart attribute analysis.

display_disk_info(): Displays all collected information in a formatted tab

```
Device: /HD1
Mountpoint: /mnt/data
File System: ntfs
Total Size: 11169.28 GB
Used: 7000.26 GB
Free: 4169.02 GB
Usage: 70%
SMART Status: ❌ Failed
SMART Attributes:
──────────────────────────────────────────────────────
ID# ATTRIBUTE_NAME          FLAG     VALUE WORST THRESH TYPE      UPDATED  WHEN_FAILED RAW_VALUE
  1 Raw_Read_Error_Rate     0x002f   100   099   051    Pre-fail  Always       -       200
  5 Reallocated_Sector_Ct   0x0033   001   001   010    Pre-fail  Always   FAILING_NOW  500
  9 Power_On_Hours          0x0032   050   050   000    Old_age   Always       -       40000
 10 Spin_Retry_Count        0x0032   098   098   097    Old_age   Always       -       2
```

Fig.9. Script results.

Second, the script transfers the necessary data for further analysis. And you can see the results of the script in Figure 9. These forensic techniques can also be repurposed for integrity monitoring in smart grid systems. By embedding such scripts into substation management software, it is possible to routinely check device health, verify firmware integrity, or detect early signs of disk degradation — especially in environments lacking traditional IT infrastructure.

Table 1. Comparison table of script and utility CrystalDiskInfo

Feature	Python Script	CrystalDiskInfo
Platform	Cross-platform (Linux, macOS, Windows)	Windows only
Disk Usage Info	Yes (psutil)	Yes
SMART Health Check	Yes (via smartctl)	Yes
Detailed SMART Data	Yes (smartctl -A)	Yes
GUI Support	No (CLI only)	Yes (Graphical Interface)
Real-time Monitoring	No (Runs on-demand)	Yes
Alerts & Notifications	No	Yes
Customization	Yes (Editable Python script)	Limited
Dependency	smartctl, psutil	Standalone EXE
Installation	Python + Dependencies	Simple EXE Install
Lightweight	Yes (Minimal overhead)	Slightly heavier

Table 1 shows that the script is not inferior to the well-known utility - CrystalDiskInfo, and in some places even surpasses it. First, the script solution will not save data, which is the main goal of forensic experts. Accordingly, there will be no monitoring and no risks.

After that we select the necessary disk as the cloning source (Fig. 10.)

While creating a copy, the process should be continuously monitored, as sounds and drive freezes may occur. (Fig. 11)

R-Studio is a program that is used for data recovery. The program is designed to work with corrupted or deleted files and partitions.

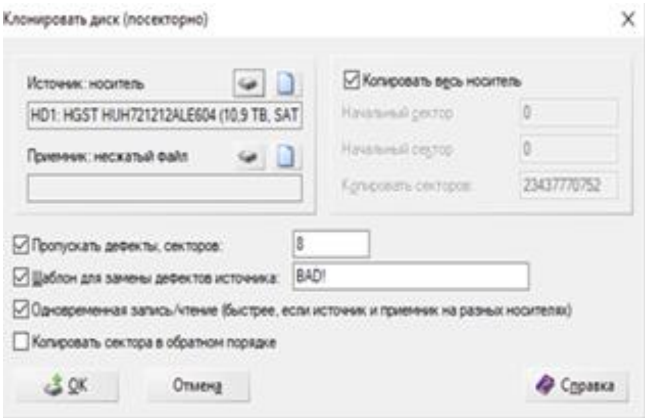


Fig.10. Disk for cloning.

Figure 12 shows an example of metadata. It should also be considered that in some cases the necessary file system metadata may no longer exist. If the files have been deleted, which is quite common in digital forensics, a quick analysis using various utilities should be prioritized. This involves scanning key structures (MFT, Index, Logfile) rather than performing a full partition scan.

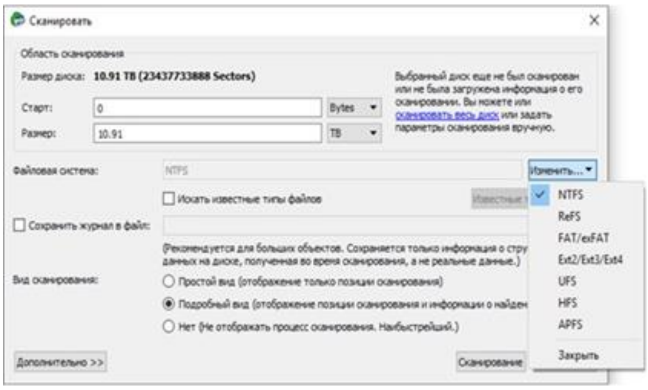


Fig.11. Scanning interface with file system selection options.

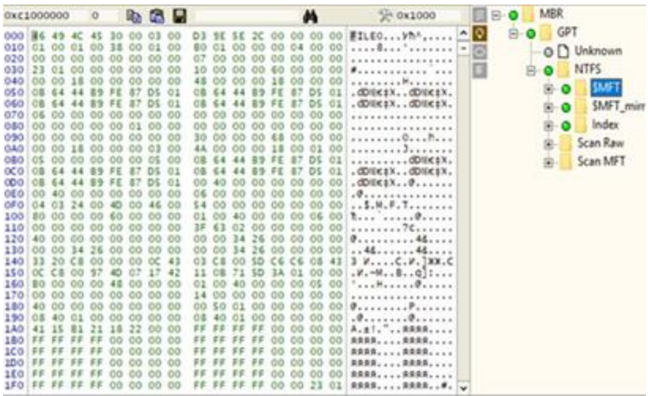


Fig.12. Example of metadata.

In such situations, searching for specific files using regular expressions will be required. For example, Figure 13 shows regular expressions for JPG files.

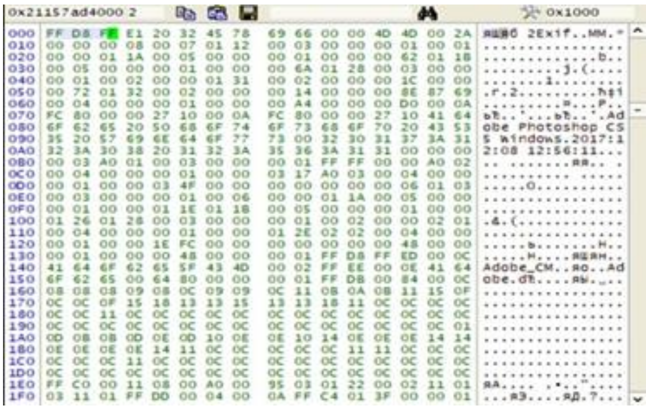


Fig.13. Regular expressions for JPG files.

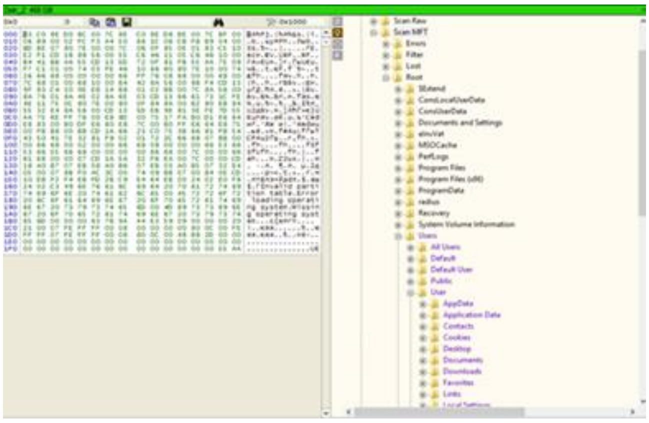


Fig. 14. Deleted data.

In Figure 14, deleted data is highlighted in purple.

This new approach to recovering data from failed hard drives represents a significant step forward in the field of data recovery. By combining existing software and hardware methods, the likelihood of successful recovery can be increased, and the process can be optimized. The integration of an automated ESR module for acoustic fault detection further enhances early diagnostics, enabling the identification of mechanical issues before physical intervention is required. This is especially important in environments where data loss can have serious consequences for users' professional and personal lives. By combining software and hardware methods, it enhances recovery efficiency and system stability in the face of technological changes and cybersecurity threats. The proposed methods are expected to find wide application and improve data recovery quality in the future.

The use of machine learning in ESR-based diagnostics also represents a computational method with direct implications for maintainability and automated fault detection in smart grid environments, where reliability and real-time response are very important.

5. Conclusions

This study introduces a novel chip-off method for data extraction from hard disk drives (HDDs), designed to address the challenges of recovering data from devices with physical or logical damage, particularly in forensic contexts. The proposed approach refines traditional chip-off techniques by incorporating precise mechanical procedures and a comprehensive recovery workflow, enabling access to raw memory chips even when conventional software-based methods fail. By leveraging specialized tools for head assembly replacement and controller board transplantation, this method ensures data integrity during extraction, offering a robust solution for retrieving critical information from damaged HDDs. The integration of an Environmental Sound Recognition (ESR) module further enhances the diagnostic phase, providing an automated, non-invasive assessment of mechanical faults through acoustic signatures, which informs and optimizes the subsequent chip-off process.

The primary contribution of this work lies in the development of an advanced chip-off technique that achieves higher success rates in data recovery by tailoring the extraction process to the specific condition of the HDD. Experimental evaluations demonstrate that this method reduces the risk of further damage during disassembly and improves recovery efficiency, with the acoustic diagnostics playing a supportive role in identifying fault types early in the workflow. The creation of a structured HDD acoustic dataset, combined with machine learning techniques such as Support Vector Machines (SVMs), complements the physical extraction by classifying audio features like Mel-Frequency Cepstral Coefficients (MFCCs) and Mel spectrograms, thus streamlining the diagnostic process.

Despite these advancements, the approach faces certain limitations. The chip-off method requires skilled technicians and controlled environments, such as cleanrooms, which may restrict its widespread adoption. For example, many small-scale forensic teams or organizations in resource-limited regions may lack access to cleanroom facilities and specialized personnel making it difficult to apply the method consistently. This limits its practical deployment outside specialized labs or high-end forensic centers. Additionally, the accuracy of the ESR module depends on the quality of audio recordings and the diversity of the training dataset, with background noise potentially impacting its reliability in real-world settings. These challenges highlight areas where the method's accessibility and robustness could be further improved.

From an IoT security perspective, the methodology presented here offers vital capabilities for investigating and recovering data from compromised or physically damaged edge devices. As smart homes, industrial systems, and critical infrastructures increasingly rely on distributed IoT nodes with onboard storage, ensuring the recoverability and integrity of such data becomes a cornerstone of digital trust and forensic readiness.

Looking ahead, future research could enhance the chip-off technique by developing more automated tools to reduce the dependency on manual expertise, thereby broadening its practical applicability. Refining the acoustic fault detection system with advanced noise reduction techniques and a more extensive dataset encompassing diverse HDD models and failure scenarios could also improve diagnostic precision. Exploring the adaptation of this method to other storage devices, such as solid-state drives (SSDs), presents an exciting opportunity to expand its scope. Ultimately, this work advances digital forensics and data recovery by offering a condition-aware, efficient chip-off solution, with the potential to evolve into a more scalable and automated approach for addressing data loss in an increasingly digital world.

By addressing the intersection of digital forensics, smart device maintenance, and acoustic signal processing, the proposed system reinforces the sustainability and resilience of data infrastructures that underpin smart homes, smart factories, and other components of the smart grid.

Acknowledgement

This study was carried out with the financial support of the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan under Contract №388/PTF-24-26 dated 01.10.2024 under the scientific project IRN BR24993232 “Development of innovative technologies for conducting digital forensic investigations using intelligent software-hardware complexes”.

References

[1] M. Gül and E. Kugu, “A survey on anti-forensics techniques,” in Proc. International Artificial Intelligence and Data Processing Symposium (IDAP), 2017.

[2] J.-P. Van Belle, “Anti-Forensics: A Practitioner Perspective,” International Journal of Cyber-Security and Digital Forensics, vol. 4, no. 2, pp. 390–403, 2015, doi: 10.17781/p001593.

[3] ABC News, “San Bernardino shooters tried to destroy phones, hard drives, sources say,” ABC News. [Online]. Available: https://abcnews.go.com/US/san-bernardino-shooters-destroy-phones-hard-drives-sources/story?id=35570286 [Accessed: Mar. 27, 2025].

[4] P. Cruickshank, “Discarded laptop yields revelations on network behind Brussels, Paris attacks,” CNN, Jan. 25, 2017. [Online]. Available: https://edition.cnn.com/2017/01/24/europe/brussels-laptop-revelations/index.html [Accessed: Mar. 27, 2025].

[5] EaseUS, Data Recovery E-Book V1.5. [Online]. Available: http://www.easeus.com [Accessed: Mar. 27, 2025].

[6] A. Silberschatz, P. B. Galvin, and G. Gagne, Operating System Concepts, 6th ed., Wiley, 2003, ch. Secondary Storage Structure.

[7] A. S. Tanenbaum and H. Bos, Modern Operating Systems, 5th ed. Harlow: Pearson Education, 2022.

[8] B. P. Battula, B. K. Rani, R. S. Prasad, and T. Sudha, “Techniques in computer forensics: A recovery perspective,” ResearchGate Journal, vol. 3, no. 2, May 2009. [Online]. Available: https://www.researchgate.net/publication/xxxxxx [Accessed: Mar. 27, 2025].

[9] A. Verducci, “How to absolutely, positively destroy your data,” Popular Mechanics, Feb. 2007. [Online]. Available: http://www.popularmechanics.com/technology/howto/4212242.html [Accessed: Mar. 27, 2025].

[10] R. Xu, X. Wang, and J. Wu, “Classification based hard disk drive failure prediction: Methodologies, performance evaluation and comparison,” in Proc. 2022 IEEE 18th Int. Conf. Automation Science and Engineering (CASE), Aug. 2022, pp. 189–195, doi: 10.1109/case49997.2022.9926720.

[11] TopTenReviews, “Top Ten Reviews,” Feb. 26, 2025. [Online]. Available: https://www.toptenreviews.com/money/business [Accessed: Mar. 27, 2025].

[12] A. Mesaros, T. Heittola, and T. Virtanen, “Metrics for polyphonic sound event detection,” in Proc. 2016 IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), 2016, pp. 351–355.

[13] D. Barchiesi, D. Giannoulis, D. Stowell, and M. D. Plumbley, “Acoustic scene classification: Classifying environments from the sounds they produce,” IEEE Signal Processing Magazine, vol. 32, no. 3, pp. 16–34, 2015.

[14] K. J. Piczak, “ESC: Dataset for environmental sound classification,” in Proc. 23rd ACM Int. Conf. Multimedia, 2015, pp. 1015–1018.

[15] J. Salamon and J. P. Bello, “Deep convolutional neural networks and data augmentation for environmental sound classification,” IEEE Signal Processing Letters, vol. 24, no. 3, pp. 279–283, 2017.

[16] K. J. Piczak, “Environmental sound classification with convolutional neural networks,” in Proc. 2015 IEEE 25th Int. Workshop Machine Learning for Signal Processing (MLSP), 2015, pp. 1–6.

[17] S. Schneider, A. Baevski, R. Collobert, M. Auli, and A. Mohamed, “wav2vec 2.0: A framework for self-supervised learning of speech representations,” in Proc. 19th Annual Conf. Int. Speech Communication Association (Interspeech), 2021, pp. 3652–3656.

[18] S. Hershey, S. Chaudhuri, D. Ellis, J. Gemmeke, A. Jansen, R. Moore, M. Plakal, D. Platt, R. Saurous, B. Seybold, M. Slaney, R. Weiss, and K. Wilson, “CNN architectures for large-scale audio classification,” in Proc. 2017 IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), 2017, pp. 131–135.

[19] Y. Su, Y. Yang, X. Liu, Q. Zhong, J. Liu, and Z. Liu, “Environmental sound classification with large-scale weakly labeled data,” in Proc. Detection and Classification of Acoustic Scenes and Events Workshop (DCASE2019), 2019, pp. 123–127.

[20] Q. Kong, Y. Cao, T. Iqbal, Y. Wang, W. Wang, and M. Plumbley, “PANNs: Large-scale pretrained audio neural networks for audio pattern recognition,” IEEE/ACM Trans. Audio, Speech, and Language Processing, vol. 28, pp. 2880–2894, 2020.

[21] F. Munir, J. Li, and Z. Yu, “Towards generalized environmental sound classification,” IEEE Access, vol. 9, pp. 72053–72065, 2021.

[22] K. J. Piczak, “ESC-50 Dataset,” GitHub Repository. [Online]. Available: https://github.com/karolpiczak/ESC-50 [Accessed: Mar. 27, 2025].

[23] L. Rzayeva, A. Imanberdi, I. Opirskyy, O. Harasymchuk, and G. Abitova, “Analysis of technical features of data encryption implementation on SD cards in the Android system,” Scientific Journal of Astana IT University, pp. 157–171, 2025.

[24] S. A. Shaikh, S. R. Biradar, and S. S. Manvi, “Cyber physical system security for smart grid: A

comprehensive survey," International Journal of Smart Grid, vol. 6, no. 1, pp. 20–32, Mar. 2022.

[25] K. Sharma and M. Bhatia, "Machine learning-based intrusion detection systems for smart grids: A review," International Journal of Smart Grid, vol. 5, no. 4, pp. 134–140, Dec. 2021.

[26] M. Alshahrani, M. Younis, and M. A. AlZain, "Cybersecurity challenges in smart grids and their mitigation using AI," International Journal of Smart Grid, vol. 4, no. 3, pp. 85–94, Sep. 2020.