COLLABORATIVE INTEGRITY AUDITING FOR SECURE AND COMPLIANT MULTI-CLOUD DATA MANAGEMENT

Robert Gordon University

09 APRIL 2025
Word Count: 2,999

Group Name: Synergy
Kenneth Boamah Adarkwah [redacted]
Donna Naadu Botchway [redacted]
Susan Ntim[redacted]

Table of Contents

Table of Figures

1.0 Introduction

1.1 Problem Statement

Organizations are rapidly adopting multi-cloud strategies to enhance flexibility, reliability, and performance. However, multi cloud service providers (CSPs) are faced with challenges related to data privacy and integrity. Different security policies, compliance rules, and protection methods create risks, like unauthorized access, data breaches, and integrity issues (Nair et al., 2024).

Reliance on third party auditors (TPAs) and high computational overhead impact efficiency and trust using traditional integrity auditing. Multi-replica storage across CSPs to enhance security and resilience is underutilized by existing solutions (Li, Chu and Hu, 2023). Addressing these challenges is key to building trust in multi-cloud environments.

To build trust in multi-cloud environments, addressing these challenges is key. This study focuses on privacy and integrity risks in multi-cloud systems, exploring collaborative integrity auditing (CIA) as a lightweight and decentralized approach to enhance security. It will evaluate how CIA can be integrated with encryption, compliance frameworks like GDPR and zero-trust models to build a consistent and trustworthy multi-cloud security model (Li, Chu and Hu, 2023; Wang et al., 2023; Liu et al., 2024).

1.2 Motivation and Research Gaps

With an increase in multi-cloud adoption, data integrity and privacy remain a challenge because of compliance gaps, varying security policies and risks like how rarely accessed data is deleted by CSPs. Following existing studies, Li, Chu, and Hu (2023), Liu et al. (2023), Nair et al. (2023) and Wang et al. (2023) proposed solutions including collaborative integrity auditing, blockchain sharding, data management framework, GDPR compliance and emerging technologies resulting in fragmented solutions. These rely on third-party auditors, introducing trust, scalability, high computational costs, and key management issues, while CSP collusion and fairness concerns persist.

This study covers Collaborative Integrity Auditing (CIA) as a decentralized lightweight solution by integrating zero-trust, encryption and GDPR compliance. This increases trust and security, reduces computational overhead and supports scalable, privacy preserving multi-cloud environments. The goal is to establish a unified security model, addressing the limitations of existing approaches.

1.3 Research Hypotheses

A unified data governance framework integrating collaborative auditing, sharded blockchain, and fair exchange protocols within a zero-trust model improves integrity, security, scalability, and compliance in multi-cloud environments.

The research addresses the following to validate the hypotheses.

Q1. How can a data governance framework improve integrity, compliance, and secure sharing while addressing fragmentation in multi-cloud environments?

Q2. Can CSPs audit each other without third parties to improve security and decrease computational costs using hash functions?

Q3. How can fair exchange, encryption and zero-trust techniques, guarantee secure data sharing and integrity audits without collusion in multi-cloud environments?

1.4 Research Aims and Objectives

Through an integrated and scalable governance framework, this study aims to improve trust in multi-cloud environments by enhancing data compliance, integrity and security (Li, Chu and Hu, 2023; Wang et al., 2023; Liu et al., 2024).

The objectives of the study are:

- Eliminate the dependency on TPAs and improve dispute resolution by allowing CSPs to audit each other.
- Decrease high computational overhead and increase performance by developing lightweight cryptographic techniques.
- Improve efficient data storage and retrieval by creating a flexible framework with user-defined block sizes.
- Ensure GDPR compliance with fair exchange and transparent integrity verification while preserving privacy.
- Leverage blockchain and zero-trust principles to improve security, prevent collusion, and enforce decentralized governance for multi-cloud audits.

2.0 Critical Review of the Literature

Organizations adopt multi-cloud approaches for flexibility and performance, but differences in security policies raise privacy and integrity issues (Nair et al, 2024). This review critically examines key studies whose findings, limitations, and diverse methodologies provide a foundation for this study. The papers used were selected following the PRISMA 2020 framework for their credibility and relevance, addressing gaps and emerging issues through empirical evidence, current trends, and foundational theories.

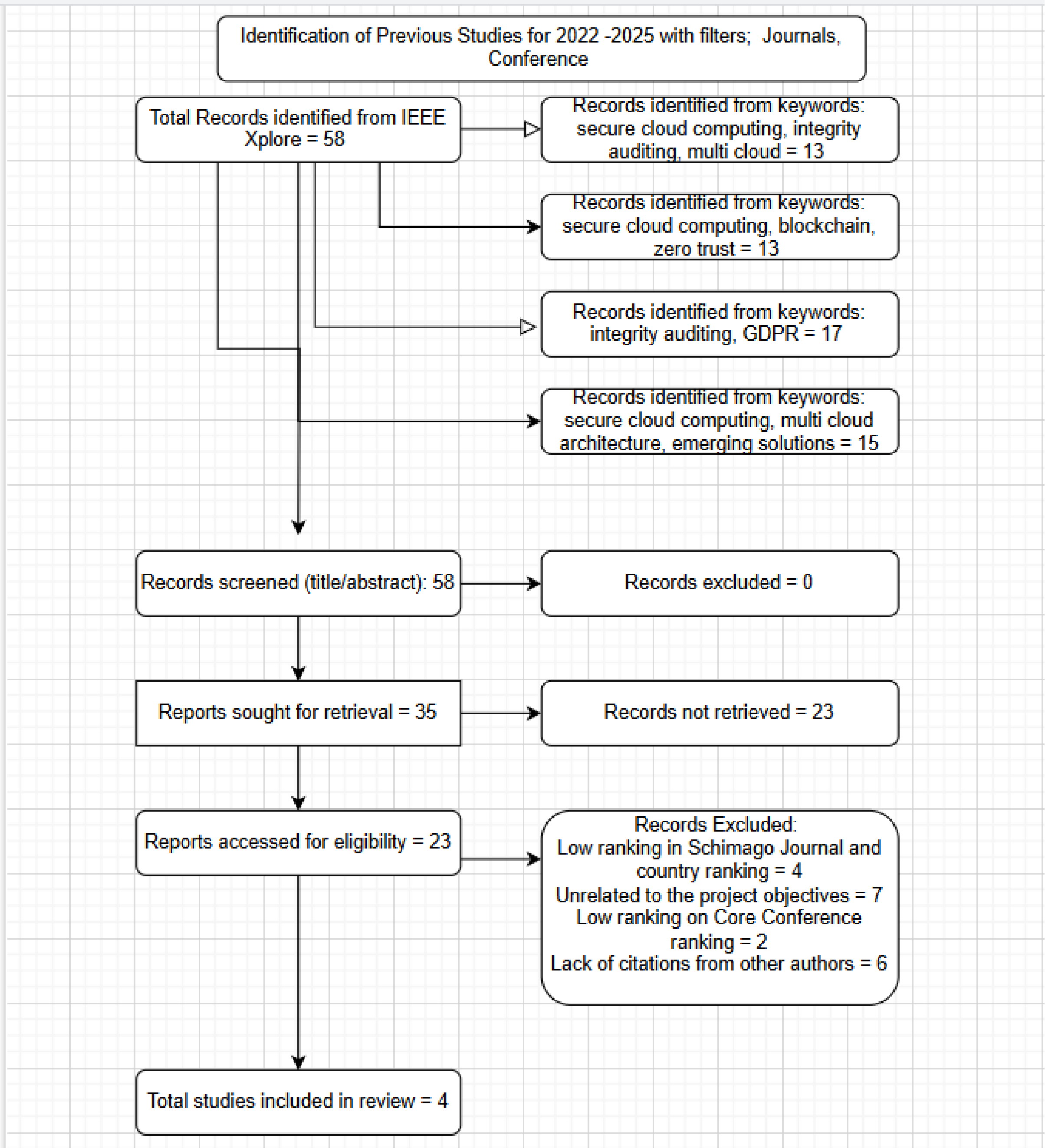

Identification of Previous Studies for 2022 -2025 with filters; Journals, Conference
Total Records identified from IEEE Xplore = 58
Records identified from keywords: secure cloud computing, integrity auditing, multi cloud = 13
Records identified from keywords: secure cloud computing, blockchain, zero trust = 13
Records identified from keywords: integrity auditing, GDPR = 17
Records identified from keywords: secure cloud computing, multi cloud architecture, emerging solutions = 15
Records screened (title/abstract): 58
Records excluded = 0
Reports sought for retrieval = 35
Records not retrieved = 23
Reports accessed for eligibility = 23
Records Excluded: Low ranking in Schimago Journal and country ranking = 4 Unrelated to the project objectives = 7 Low ranking on Core Conference ranking = 2 Lack of citations from other authors = 6
Total studies included in review = 4

Figure1: PRISMA_2020 Methodology for paper selection

Li, Chu and Hu (2023) suggest an approach to improve data integrity verification in multi-cloud storage environments. The main findings of the paper are: (i) a lightweight CIA scheme eliminates the dependence of third-party auditors (TPA) endorsing cloud service providers (CSPs) to audit each other. (ii) computational overhead reduction by removing the tag generation phase which contributes to high computational overhead in traditional auditing models (iii) flexibility as data owners customize block sizes (160-bit minimum) according to their needs without compromising security. (iv) CIA eliminates CSP manipulation using the SecNego and Proof-Distr stages without compromising data integrity and efficiency in multi CSP scenarios. However, this study assumes the independence and honesty of CSPs. This causes the scheme to be at risk of manipulation if there is collusion or sharing of sensitive information, compromising data integrity. When there is uncertainty of trust, the scheme is not fully applicable. Again, the scheme fails to address data dynamics challenges where data update and deletion is frequent, limiting users with highly dynamic data. The study is limited to a sizable replicas and less complex data structures which may compromise scalability since computational overhead may still be high. Finally, the lack of a comprehensive security analysis leaves vulnerabilities unaddressed making it less desirable in real-world applications.

Nair et al (2023) provide a comprehensive overview of the current state of multi cloud architecture by identifying key challenges and best practices of multi-cloud security. They highlight best practices and how to develop strategies for security challenges in multi-cloud environments. The paper's methodology includes extensive incorporation of various research sources leading to a diverse approach to enhance security with the integration of machine learning and big data. The study lacks quantitative data and focuses on specific industries limiting the application of its conclusions to a larger audience. Other limitations include insufficient exploration of emerging technologies, identity access management issues and challenges in compliance and integration.

Wang et al (2024) outlines practical implementation of cross-domain data sharing in a zero-trust cloud-edge-end environment. The paper explores different methods such as plaintext checkable encryption scheme ensuring data integrity without decrypting the data, multi-domain architecture using blockchain sharding to distribute data across multiple nodes and cross-domain data sharing scheme under both partial and zero trust amongst entities. The findings of the study include (i) a framework for data sharing in instances where entities do not inherently trust each other. (ii) a detailed theoretical analysis of proposed protocols suggesting efficient policy registration and execution with real-world applications. (iv) a comprehensive security analysis demonstrating the protocols against threats without compromising data integrity or confidentiality. However, large-scale implementations face scalability issues due to limitations of traditional schemes with increasing participating nodes. In real-world scenarios, assuming messages between honest hosts in a partial synchronous communication model arrive in a specified time may affect performance of the proposed protocols. Although the paper analyses security, attackers can explore new and sophisticated techniques to attack the protocols limiting their robustness against threats. Practical implementation and broader applicability in the real-world scenarios need to be addressed.

Liu et al (2023) focus on integrating blockchain with modern auditing techniques whilst ensuring data integrity and compliance with GDPR. The paper presents a provable data possession (PDP) scheme which reduces bandwidth by checking data integrity and semantic consistency concurrently, using smart contracts to encode data usage policies and promote GDPR compliance. The findings of the paper emphasize on a framework which enhances and addresses data management practices and its security concerns. The framework allows data owners to verify the compliance and integrity of their data in cloud environments under GDPR. The framework reduces computational costs and enhances security by simplifying certificate management and making a secure channel optional. The PDP scheme struggles to verify data integrity and semantic consistency simultaneously which is crucial for GDPR compliance and may compromise the auditing process effectiveness. This framework also relies on the honesty of involved entities to generate correct proof. The implementation is quite complex and expensive, limiting real-world applications and deterring low budget organizations. Finally, bilinear pairing operations for verification are time-consuming leading to performance bottlenecks.

In conclusion, all reviewed papers face challenges in multi-cloud environments including high computational cost, data security concerns, scalability issues and complexities of proposed solutions. To address these challenges, the study focuses on integrating blockchain, encryption, zero-trust frameworks and GDPR compliance with CIA to improve security, scalability and data integrity.

3.0 Legal, Ethical, Social and Professional Issues

3.1. Legal Implications

This study aligns with legal frameworks, particularly the GDPR following its ‘right to be forgotten’ data minimization principles. It uses blockchain for secure transactions and accountability while protecting data privacy (Wang et al., 2023). Simulated transactions test compliance and security without legal issues. The study ensures fair data use with tools like XACML and SecKit (Nair et al., 2024) following the NIST Cybersecurity Framework. It uses continuous monitoring and immutable logs to detect threats, maintain accountability, and breach recovery (Wang et al., 2023; Nair et al., 2024).

3.2. Ethical Considerations

This study guarantees integrity by using synthetic datasets instead of real user data. It eliminates the risks associated with handling real-life datasets. The use of blockchain-based integrity verification and zero-trust security models prevents unfair advantages and data manipulation (Wang et al., 2023; Nair et al., 2024). It guarantees fairness by reducing collusion in auditing systems. Using synthetic data allows thorough security evaluations without exposing sensitive data, ensuring the findings are valid to real-world scenarios.

3.3. Social Impact

This study improves system accessibility, scalability, and reliability in addressing the increase of data-sharing participants. Sharded blockchain technology distributes workloads across multiple nodes ensuring scalability (Liu et al., 2023). By combining with the zero-trust architecture, trust is improved in multi-cloud environments (Li, Chu, & Hu, 2023). The study ensures data integrity, compliance, accountability, and privacy in data-critical sectors like healthcare and supply chain management while reducing auditing costs. However, it is important to develop the project responsibly to avoid unfair advantages, exclusion of marginalized groups, and address concerns like implementation costs, technical challenges, data loss, and inclusivity.

3.4. Professional Issues

This study requires expertise in cloud computing, blockchain, cryptography, and data governance, which may highlight skill gaps for professionals in these fields. It presents opportunities for professional development by promoting fair exchange protocols and improving collaborative auditing practices in cloud security. Professionals are encouraged to follow best practices and ensure compliance aligning with industry standards such as GDPR and NIST. This helps them grow in data security and cloud management. Furthermore, it encourages transparency, fosters ethical decision-making, and accountability among professionals.

4.0 Research Methodology

4.1 Data Collection and Preparation

This study uses simulated and synthetic datasets to ensure ethical compliance, privacy, and controlled testing. Simulated cloud environments and blockchain networks evaluate security, integrity, and compliance mechanisms.

The Collaborative Integrity Auditing scheme verifies file integrity through hashing and encrypted replicas (Li, Chu, and Hu, 2023). Liu et al. (2024) use synthetic datasets to mimic zero-trust cloud sharing, including transaction logs and audit trails. Blockchain-based GDPR compliance testing is simulated using compliance rules, and metadata (Wang et al., 2023).

These datasets allow for precise security evaluations while eliminating privacy risks and ensuring reproducibility.

4.2 Research Methods & Strategies

This study follows a design science methodology to implement innovative security technologies to support continuous cloud integrity and compliance. The study ensures a structured yet flexible approach to problem-solving by combining theoretical modelling, simulations, and comparative analysis,

Inspired by Li, Chu, and Hu (2023), CIA enables CSPs to audit each other, reducing reliance on third-party auditors. Files are split into blocks, hashed, and distributed across CSPs, which verify data integrity through challenge-response exchanges.

Wang et al. (2023), highlights a blockchain-based framework to ensure GDPR compliance using a provable data possession (PDP) scheme which stores hashed metadata on the blockchain, preventing tampering. To test compliance and security, transactions are simulated.

To improve scalability and security in zero-trust cloud sharing, Liu et al. (2024) proposed a sharded blockchain model. It mimics cloud-edge environments, testing cross-domain consistency, policy registration, and data-sharing security.

Security models are tested under various conditions, measuring latency, integrity verification rates, and compliance success. These provide insights into balancing security, efficiency, and compliance in real-world scenarios.

4.3 Tools, Technologies, and Performance Metrics

A range of tools, including cloud testbeds, blockchain frameworks, cryptographic libraries and programming languages, are adopted for the proposed solution.

Following Li, Chu, and Hu (2023), AWS, Azure, and Google Cloud Platform simulate zero-trust cloud sharing. Hyperledger Fabric or Ethereum (Liu et al., 2023) are adopted for audit logging.

Building on Li, Chu, and Hu (2023), for efficiency the C language with GMP handles cryptographic arithmetic operations. PBC for cryptographic functions and SHA-256 for secure hashing (Liu et al., 2023) tested on Ubuntu.

For GDPR-compliant blockchain implementation, Java with My Eclipse IDE for development, JPBC for cryptographic functions and Quorum for compliance auditing (Wang et al., 2023). The security and efficiency of the framework will be measured using four metrics: throughput for integrity audits, and latency of for verification response time. Unbiased auditing and non-collusion among CSPs will be measured through the fairness of the framework. Finally, to ensure the resilience of the solutions against attacks like data tampering and unauthorized access, security will be measured.

The tools and metrics will ensure relevant testing, optimization and validation of the proposed security framework.

4.4 Datasets Available

Synthetic datasets are generated across all three solutions to effectively validate collaborative auditing schemes, blockchain-based GDPR compliance, and zero-trust architectures to safeguard sensitive information.

Synthetic datasets are replicated for CSP interactions to enhance privacy and compliance in the context of zero-trust cloud sharing. These data sets include access policies, disputes and transaction (Liu et al., 2023).

Data blocks and test replicas are simulated to represent data sizes and challenge parameters for integrity without compromising efficiency and security in Collaborative Integrity Auditing (Li, Chu and Hu, 2023).

Lastly audit logs, user metadata, and access control policies are simulated for blockchain-based GDPR to assess compliance and supervise rules and assess associated costs (Wang et al., 2023).

These synthesized datasets ensure a fair evaluation, boosting performance and resilience against attacks by comparing traditional integrity auditing methods and the proposed decentralized approach. To assess the efficiency and security trade-offs of the proposed framework key parameters like data size, policies and replication factors can be used.

5.0 Feasibility, Significance, and Innovation

Using simulated data to create realistic testing conditions makes this study highly feasible while ensuring ethical compliance. Hyperledger Fabric and Ethereum blockchain technologies are used for protocol review whilst testbeds like AWS, Azure and Google Cloud provide scalable experimentation environments. SHA-256 and GMP cryptographic technologies ensure safe calculations. Applications involving extensive data exchange use sharded-blockchain technology as it increases scalability. C and Python are used for implementation, tested on Ubuntu and published on GitHub (Li, Chu and Hu, 2023; Wang et al., 2023; Liu et al., 2024).

The research promotes collaborative integrity auditing and fair exchange protocols contributing to cloud security advancement in academia. The study reduces auditing costs, enhances trust, improves security and ensures regulatory compliance in the industry. For society, it enforces data integrity and accountability while promoting privacy-preserving data sharing in sectors like supply chain management and healthcare. The study fosters the use of secure cloud computing in a zero-trust model by addressing conflicts and discouraging malicious actions.

6.0 Timeline

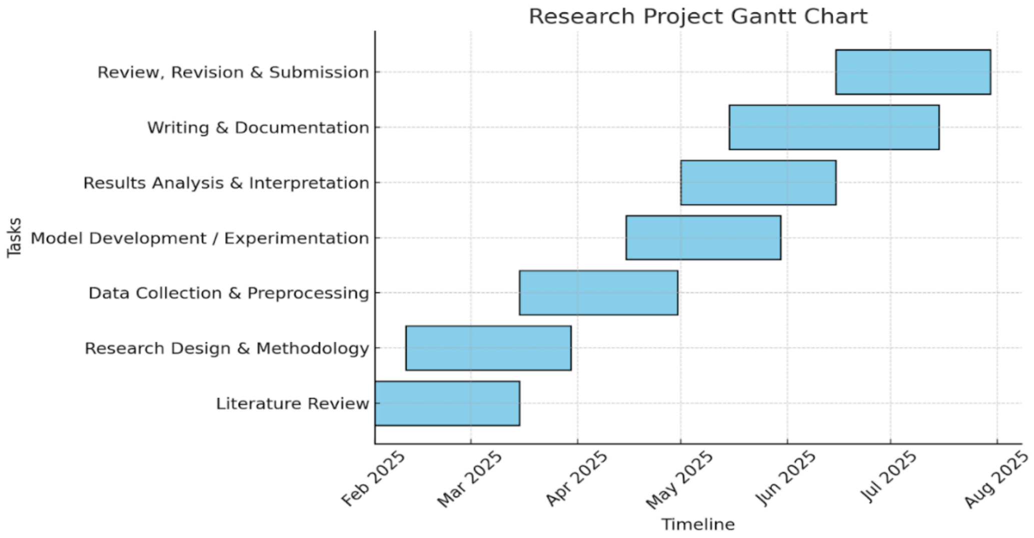


Figure 2: Timeline.

The study timeline spans from February to August 2025 as shown in the Gantt chart above. From February to March, the literature review commences to examine previous studies, including their findings, research gaps and methodologies. The research design and methodology phase commence from March to April using the design research methodology. Next, from April to May we collect relevant data (create synthetic data), clean and preprocess the data. Then from May to June, we build and test the proposed system.

During the result analysis and interpretation phase, research findings are interpreted from May to June. Following the results interpretation, the writing and documentation phase takes place. The last phase, review, revision and submission, runs from June and July to ensure all necessary revisions are made. To meet the project deadlines and increase efficiency, some duties are scheduled to overlap.

7.0 Conclusions from Statistical Analysis

This study used statistical methods in R to explore clinical data to identify what factors affect glaucoma (GRS) and cataract (CAT_Score) risk. These included Shapiro-Wilk test to check data normality, Pearson's and Spearman's correlations for relationships, and other distribution-based.

The dataset had parametric and non-parametric variables. Intraocular pressure (IOP) and corneal thickness followed a normal distribution, while age, pupil diameter, refractive error, and visual acuity did not. GRS was normally distributed, while CAT_Score was not.

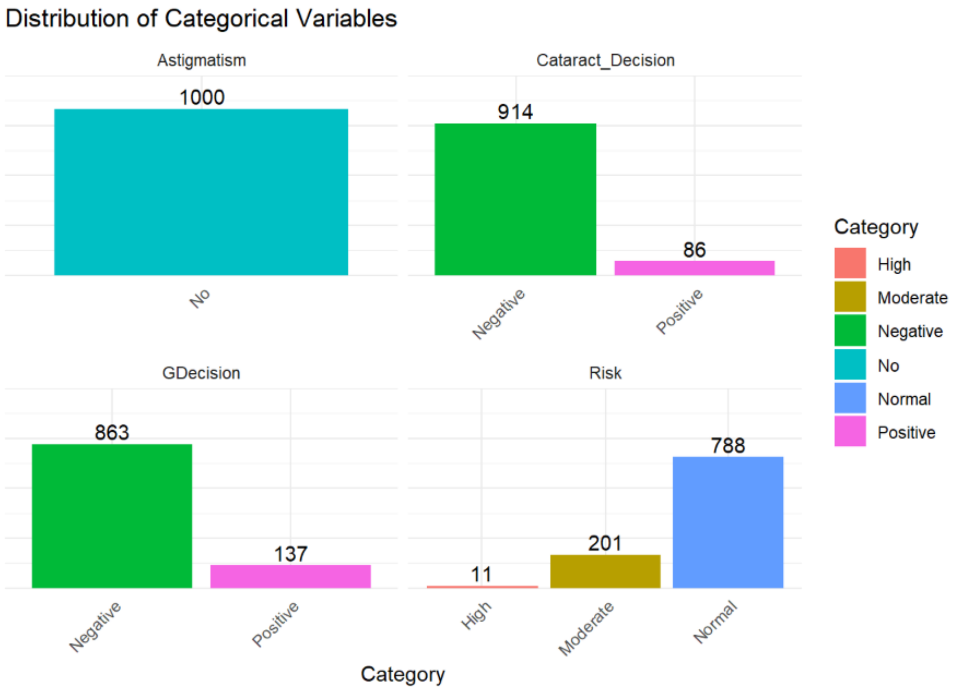


Figure 3: Categorical Distribution.

Figure above shows the distribution of categorical data. Astigmatism was excluded because it had no variability as all patients had 'No'. Figure 4, a stacked bar chart shows risk distribution across genders, revealing no major gender-based differences.

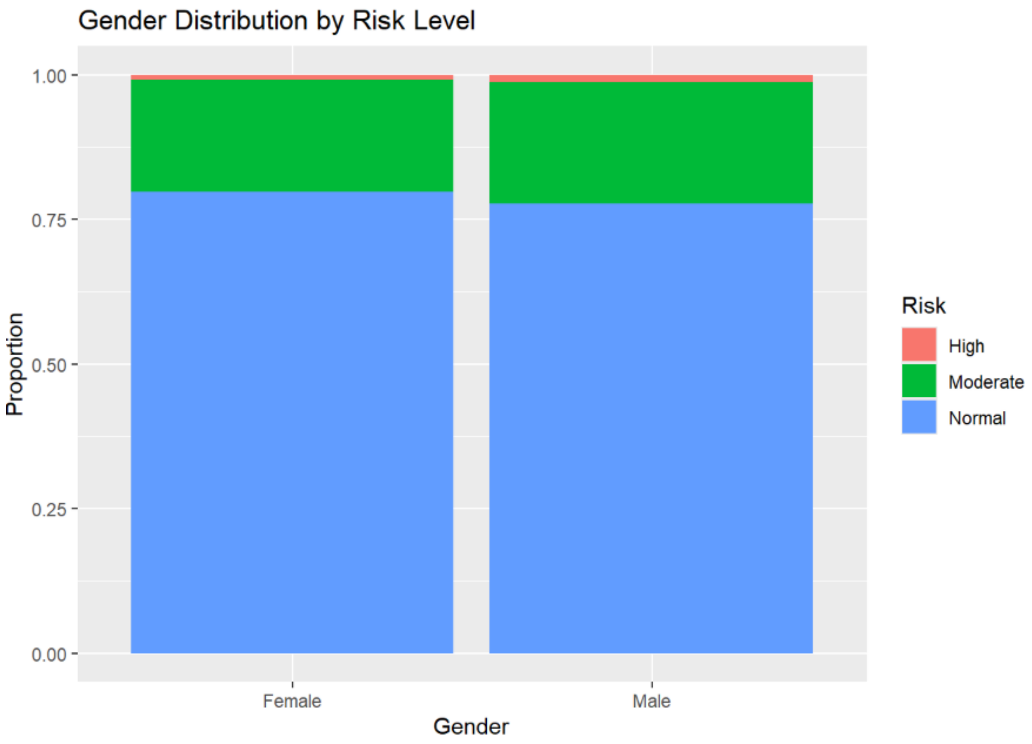


Figure 4: Gender Distribution by Risk.

Test selection followed data type and distribution: ANOVA was used when both variables were normally distributed (like GRS vs. IOP); Kruskal-Wallis was used when one was non-normal (like GRS vs. age); Wilcoxon was used when both were non-parametric (like CAT_Score vs. age). Chi-square tests were used for categorical variables like Risk, GDecision, and Cataract_Decision. Figures below present the correlation results, test values, and conclusions.

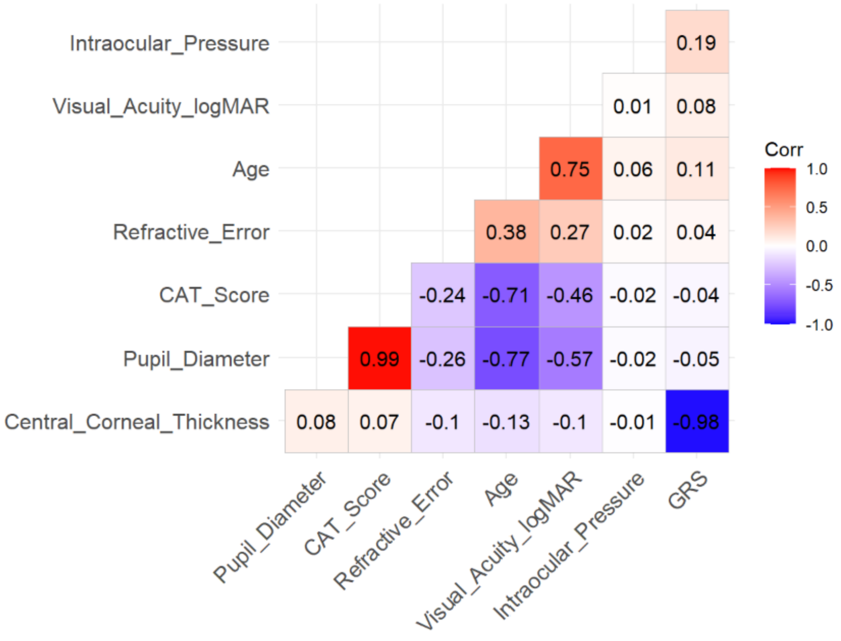
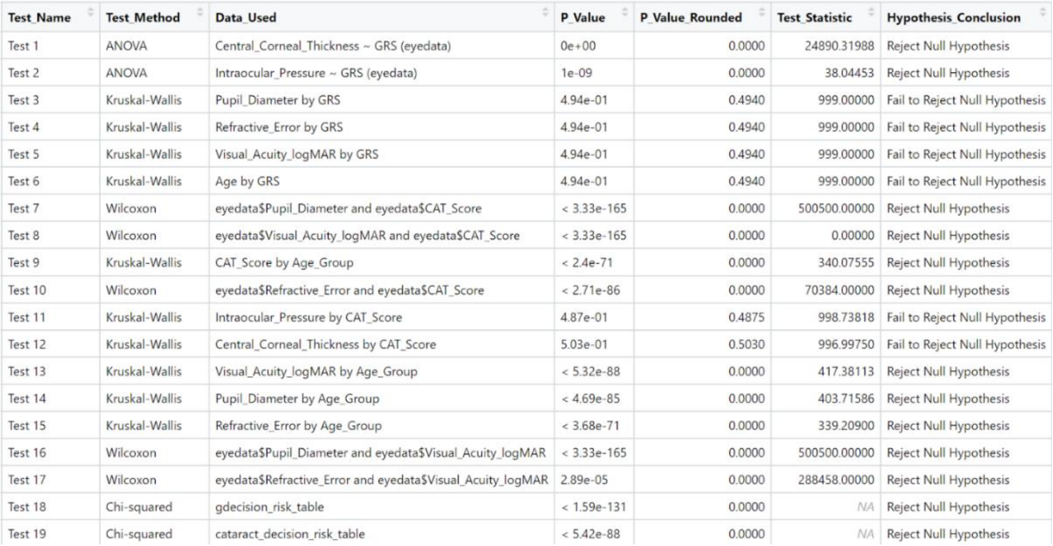


Figure 5: Correlation Matrix.

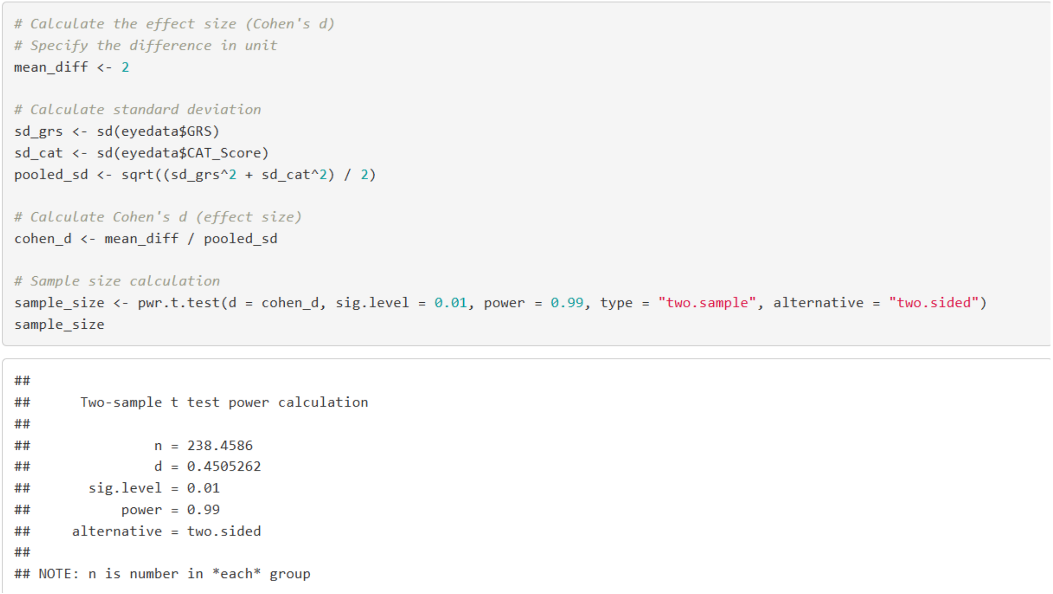


Test_Name	Test_Method	Data_Used	P_Value	P_Value_Rounded	Test_Statistic	Hypothesis_Conclusion
Test 1	ANOVA	Central_Corneal_Thickness ~ GRS (eyedata)	0e+00	0.0000	24890.31988	Reject Null Hypothesis
Test 2	ANOVA	Intraocular_Pressure ~ GRS (eyedata)	1e-09	0.0000	38.04453	Reject Null Hypothesis
Test 3	Kruskal-Wallis	Pupil_Diameter by GRS	4.94e-01	0.4940	999.00000	Fail to Reject Null Hypothesis
Test 4	Kruskal-Wallis	Refractive_Error by GRS	4.94e-01	0.4940	999.00000	Fail to Reject Null Hypothesis
Test 5	Kruskal-Wallis	Visual_Acuity_logMAR by GRS	4.94e-01	0.4940	999.00000	Fail to Reject Null Hypothesis
Test 6	Kruskal-Wallis	Age by GRS	4.94e-01	0.4940	999.00000	Fail to Reject Null Hypothesis
Test 7	Wilcoxon	eyedata$Pupil_Diameter and eyedata$CAT_Score	< 3.33e-165	0.0000	500500.00000	Reject Null Hypothesis
Test 8	Wilcoxon	eyedata$Visual_Acuity_logMAR and eyedata$CAT_Score	< 3.33e-165	0.0000	0.00000	Reject Null Hypothesis
Test 9	Kruskal-Wallis	CAT_Score by Age_Group	< 2.4e-71	0.0000	340.07555	Reject Null Hypothesis
Test 10	Wilcoxon	eyedata$Refractive_Error and eyedata$CAT_Score	< 2.71e-86	0.0000	70384.00000	Reject Null Hypothesis
Test 11	Kruskal-Wallis	Intraocular_Pressure by CAT_Score	4.87e-01	0.4875	998.73818	Fail to Reject Null Hypothesis
Test 12	Kruskal-Wallis	Central_Corneal_Thickness by CAT_Score	5.03e-01	0.5030	996.99750	Fail to Reject Null Hypothesis
Test 13	Kruskal-Wallis	Visual_Acuity_logMAR by Age_Group	< 5.32e-88	0.0000	417.38113	Reject Null Hypothesis
Test 14	Kruskal-Wallis	Pupil_Diameter by Age_Group	< 4.69e-85	0.0000	403.71586	Reject Null Hypothesis
Test 15	Kruskal-Wallis	Refractive_Error by Age_Group	< 3.68e-71	0.0000	339.20900	Reject Null Hypothesis
Test 16	Wilcoxon	eyedata$Pupil_Diameter and eyedata$Visual_Acuity_logMAR	< 3.33e-165	0.0000	500500.00000	Reject Null Hypothesis
Test 17	Wilcoxon	eyedata$Refractive_Error and eyedata$Visual_Acuity_logMAR	2.89e-05	0.0000	288458.00000	Reject Null Hypothesis
Test 18	Chi-squared	gdecision_risk_table	< 1.59e-131	0.0000	NA	Reject Null Hypothesis
Test 19	Chi-squared	cataract_decision_risk_table	< 5.42e-88	0.0000	NA	Reject Null Hypothesis

Figure 6: Statistical Test Overview.

GRS was significantly impacted by IOP and corneal thickness, supporting their known role as glaucoma risk factors (National Eye Institute, 2025). Higher eye pressure and thinner corneas were associated with increased GRS. Age, refractive error, pupil size, and visual acuity had little to no effect on GRS. In contrast, CAT_Score was affected by non-parametric variables: older patients and those with poorer visual acuity, larger pupil sizes, and higher refractive error had more severe cataracts. IOP and corneal thickness did not influence CAT_Score. Also, the Chi-square tests showed glaucoma and cataract decisions affected risk classification.

A sample size analysis to detect a 2-unit difference between GRS and CAT_Score at 1% significance and 99% power showed that 238 participants per group (477 total) are required (Figure7).



```
# Calculate the effect size (Cohen's d)
# Specify the difference in unit
mean_diff <- 2

# Calculate standard deviation
sd_grs <- sd(eyedata$GRS)
sd_cat <- sd(eyedata$CAT_Score)
pooled_sd <- sqrt((sd_grs^2 + sd_cat^2) / 2)

# Calculate Cohen's d (effect size)
cohen_d <- mean_diff / pooled_sd

# Sample size calculation
sample_size <- pwr.t.test(d = cohen_d, sig.level = 0.01, power = 0.99, type = "two.sample", alternative = "two.sided")
sample_size
```

```
##
##      Two-sample t test power calculation
##
##               n = 238.4586
##               d = 0.4505262
##       sig.level = 0.01
##           power = 0.99
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

Figure7: Sample Size Analysis.

These results align with existing research and point to the clinical value of monitoring these variables. Future work could explore genetic and lifestyle factors for better insight.

8.0 References

ChatGPT. (2025). Statistics Analysis Query. Available at: https://chatgpt.com/share/67f694c4-e654-8000-88a2-a40379241d94 (Accessed: 5 April 2025).

Li, T., Chu, J., and Hu, L. (2023) 'CIA: A Collaborative Integrity Auditing Scheme for Cloud Data with Multi-Replica on Multi-Cloud Storage Providers', *IEEE Transactions on Parallel and Distributed Systems*, 34(1), pp. 154-162. Available at: https://doi.org/10.1109/TPDS.2022.3216614

Lui, Y. et al., "Secure and Scalable Cross-Domain Data Sharing in Zero-Trust Cloud-Edge-End Environment Based on Sharding Blockchain," in IEEE Transactions on Dependable and Secure Computing, vol. 21, no. 4, pp. 2603-2618, Available at https://doi.org/10.1109/TDSC.2023.3313799

Nair, R.R., Sreevidya, D., Mohan, C.R., Banerjee, J. and Chouhan, K., 2024. Comprehensive Approaches to Securing Multi-Cloud Architectures: Best Practices and Emerging Solutions. In: 2024 7th International Conference on Contemporary Computing and Informatics (IC3I), September 2024, Volume 7, pp. 1631-1636. IEEE. Available at https://doi.org/10.1109/IC3I61595.2024.10828803

National Eye Institute (2025) Glaucoma and eye pressure. Available at: https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/glaucoma/glaucoma-and-eye-pressure (Accessed: 5 April 2025).

Wang, L., Guan, Z., Chen, Z. and Hu, M. (2023) 'Enabling integrity and compliance auditing in blockchain-based GDPR-compliant data management', IEEE Internet of Things Journal, 10(23), pp. 20955–20967. Available at https://doi.org/10.1109/JIOT.2023.3285211