World Journal of Pharmaceutical and Life Sciences

www.wjpls.org

Impact Factor: 7.409 Coden USA: WJPLA7

DEVELOPMENT OF NOVEL GENOMIC DATA ANALYSIS TECHNIQUES IN SAUDI ARABIAN MEDICAL LABORATORIES

Khaled Alshehri[1], Dr. Rawabi Z. Zahed[2], Nada Alghamdi[3], Rami M Alghamdi*[4], Gharam Y. Alruwaili[5], Abdulaziz Alzahrani[6], Mushrif M Alamari[7], Abdullah H Alharbi[8], Yassir Bin Khalifah[9], Afrah M Alenzi[10], Mubarak Al Mohammal[11], Abdulaziz Alsahli[12], Rayan H Alsubaie[13], Abdulrahman Aljeathen[14], Hamoud Alzahrani[15]

[1]Bachelor of Medicine and Surgery, Imam Mohammad Ibn Saud Islamic University, Kingdom of Saudi Arabia.
[2]PhD in Molecular Biology, Assistant professor, Department of Biological sciences, Faculty of Science, King Abdulaziz University (KAU), Jeddah, Kingdom of Saudi Arabia
[3]Medical Technology, National Guard Hospital, Kingdom of Saudi Arabia
[4]Lab Specialist, Master Student, Albaha Health Cluster, Kingdom of Saudi Arabia.
[5]Laboratory Specialist, Master Degree in Clinical Laboratory Sciences (Medical Microbiology), Department of Clinical Laboratory Sciences College of Applied Medical Sciences, Jouf University, Laboratory Specialist, Master Degree in Clinical Laboratory Sciences (Medical Microbiology) , Laboratory Department, Dr. Sulaiman Al Habib Medical Group, Riyadh, Saudi Arabia.
[6]Medical Laboratory Technician, General Administration of Medical Services (Al-Nakheel Medical Center), Kingdom of Saudi Arabia.
[7]Laboratory Technician, (General Directorate of Medical Services), Kingdom of Saudi Arabia.
[8]Laboratory Specialist, Forensic Toxicology Services Department, Kingdom of Saudi Arabia.
Laboratory tech, Riyadh Forensic Toxicology Services Department, Kingdom of Saudi Arabia
[10]Laboratory Specialist, Qassim University, Kingdom of Saudi Arabia.
[11]Bachelor of Medical Laboratory Sciences, Internal Security Forces Hospital in Dammam, Kingdom of Saudi Arabia
[12]Laboratories and Medical Technology, King Saud University, College of Applied Medical Sciences, Riyadh, Kingdom of Saudi Arabia.
[13]Medical Laboratory Specialist, Hail, Kingdom of Saudi Arabia.
[14]Master student of microbiology, Kingdom of Saudi Arabia.
[15]Laboratory Technician, King Abdullah Medical Complex, Jeddah, Saudi Arabia.

***Corresponding Author: Rami M Alghamdi**
Lab Specialist, Master Student, Albaha Health Cluster, Master Student, Albaha Health Cluster, Kingdom of Saudi Arabia.
Email ID: ralghamdi1641@stu.kau.edu.sa
DOI: https://doi.org/10.5281/zenodo.17777287

How to cite this Article:- Khaled Alshehri, Dr. Rawabi Z. Zahed, Nada Alghamdi, Rami M Alghamdi*, Gharam Y. Alruwaili, Abdulaziz Alzahrani, Mushrif M Alamari, Abdullah H Alharbi, Yassir Bin Khalifah, Afrah M Alenzi, Mubarak Al Mohammal, Abdulaziz Alsahli, Rayan H Alsubaie, Abdulrahman Aljeathen, Hamoud Alzahrani**.** (2025). Bachelor of Medicine and Surgery, Imam Mohammad Ibn Saud Islamic University, Kingdom of Saudi Arabia. World Journal of Pharmaceutical and Life Science, 11(4), 261–276.

Article Received on 04/04/2025 **Article Revised on 24/04/2025** **Article Published on 30/04/2025**

ABSTRACT

Genomic data analysis is, indeed, a transformative factor for personalized medicine and precision diagnostics. This research undertakes the study of the adoption and performance of advanced genomic analysis techniques, including AI models, in 160 medical labs across six regions in Saudi Arabia. Survey data and statistical analysis (SPSS) were employed to assess: the impact of AI on diagnostic accuracy; the assessment of execution efficiency in WGS/RNA-seq pipelines; and regional variations, governance of data, and readiness from an ethical point of view. The results demonstrated that AI methods greatly improve diagnostic accuracy ($p < 0.001$) and that deep learning models reduce execution time ($p = 0.0086$). However, inconsistencies in data sharing, privacy, and ethical approval practices call for a standardized framework. The study highlights the progress being made in Saudi Arabia toward genomics-enabled healthcare while identifying operational and regulatory issues to be addressed for the realization of the Vision 2030 project.

Index Terms- Genomic data analysis, AI in healthcare, deep learning, Saudi Arabia, medical laboratories, diagnostic accuracy, ethical approval, data privacy, personalized medicine.

I. INTRODUCTION

Background

Genomic data analysis has been a revolutionary component in medical diagnostic processes, providing a greater elucidation of the genetic cause of disease, personalized treatment therapies, and improved patient outcomes.[1],[2] Globally, advancement of genomics technologies such as Whole Genome Sequencing (WGS), RNA sequencing (RNA-seq), and CRISPR-mediated gene editing has revolutionized biomedical science and clinical practice.[3],[4] The technologies allow high-throughput and fast sequencing of genetic material equivalent to finding mutations, gene expression signatures, and epigenetic modifications relevant to complex disorders such as cancer, rare genetic disorders, and infectious diseases.[5]

Genomic data-based diagnostics offer precision medicine choices depending on a patient's unique genetic profile.[6] As examples, targeted cancer therapies, drug response prediction through pharmacogenomics, and presymptomatic diagnosis of genetic diseases are becoming available on a worldwide level.[7] The introduction of Artificial Intelligence (AI) and Machine Learning (ML) in genomic data analysis is facilitating the ability to process enormous amounts of data, identify patterns, and build predictive models for increased diagnostic accuracy and efficiency.[8],[9]

Genomic medicine is of greater profile in Saudi Arabia as part of the nation's push towards healthcare quality enhancement and promotion of personalized medicine.[10] The Saudi Human Genome Program (SHGP) that was launched in 2013 is a reflection of the country's interest in establishing genomic research and its transfer to clinical practice. With the prevalence of genetic disorders and consanguineous marriages in the country, genomic data analysis has immediate potential to enhance diagnostic efficiencies and personalized medical treatments.[11],[12] Next-generation sequencing technologies and bioinformatics tools are being integrated in Saudi medical laboratories to advance genomic diagnostics, while the integration of innovative data analysis platforms remains in early stages.[13]

Problem Statement

Despite international advances and domestic efforts, Saudi Arabian medical labs are subject to a variety of limitations in realizing genomic data analysis promise. There are significant knowledge gaps in the adoption and creation of new genomic analysis techniques that are tailored to the specific needs of Saudi patients and health care systems. The majority of the laboratories continue to use traditional data handling techniques that are time-consuming, nonscalable, and sometimes suboptimal for large-scale data generated by next-generation sequencing platforms.

Also, use of AI and ML models in genomic diagnostics is not common in Saudi Arabia because of a lack of expertise, insufficient computing resources, and a lack of local datasets available. These factors have consequences on accuracy, speed, and reproducibility of diagnostic findings, with ultimate effects extending to patient care. Variations in healthcare facilities and access to advanced genomic tools in various regions of the nation are also responsible contributors to uneven quality of diagnostics in various regions of the nation.

Another key challenge lies in data governance and ethical standards. Data sharing policies and privacy and regulation issues of genomic data are underdeveloped or applied inconsistently in Saudi medical laboratories. This discourages coordination, data integration, and large-scale studies that are necessary to drive genomic diagnostics forward. In addition, insufficient awareness and training of healthcare professionals in genomic data analysis impedes the implementation of state-of-the-art approaches.

All of these issues together emphasize the need to create and implement novel genomics data analysis techniques that are fast, accurate, and context-specific to Saudi Arabia's healthcare environment.

Significance of This Study

This study is important due to a number of factors. Firstly, it comes in response to the pressing need to update genetic data analysis in Saudi hospitals through investigation and utilization of new computing techniques such as AI and ML-based techniques. By identifying strengths and limitations of known techniques, this study has a potential to provide guideline to optimizing pipelines that improve diagnostic accuracy and minimize processing time to the ultimate good of the patient.

Second, it aids in personalized medicine efforts through more accurate genetic diagnoses, which, in the case of the population of Saudi Arabia, are especially relevant considering peculiar genetic backgrounds and prevalence of diseases. More sophisticated genomic diagnostics have the ability to allow early identification and personalized treatments for genetic disorders, cancers, and other genetically caused diseases that are common in the region.

Third, the research establishes regional disparities and infrastructure shortages and provides valuable information to policymakers, health administrators, and laboratory managers to plan and equip training programs accordingly. Through its emphasis on data sharing, privacy, and ethics, the research promotes the development of robust regulatory mechanisms to enable collaborative and responsible genomic research.

Lastly, the result of such a study will propel adoption of emerging genomics data analysis technologies in Saudi Arabia to a leading position in genomic medicine in the Middle Eastern region. Not only are national healthcare

standards improved, but it also serves to help support global effort towards a better understanding of genetics in human beings and the establishment of new diagnostic technologies.

Research Questions & Hypotheses

According to the issues raised and objectives of the study, below are the study questions and hypotheses that inform this study:

Research Question

What are the recent genomic data analysis techniques employed in medical laboratories in Saudi Arabia?

How are these approaches superior in accuracy and efficiency in the diagnosis of inherited disorders compared to the traditional methods?

What AI and machine learning models are most suitable for use in the analysis of genomic data in Saudi Arabian clinical environments?

Regional differences in usage and application of genomic analysis software by Saudi medical laboratories exist and have been observed.

What are the privacy, ethical, and data-sharing concerns that revolve around Saudi Arabian medical labs' genomic research?

Hypotheses

H1: The use of AI-based genomics analysis techniques employed in medical laboratories in Saudi Arabia improves diagnostic accuracy significantly compared to traditional methods.

H2: Genomic technologies like Whole Genome Sequencing and RNA sequencing, combined with deep learning models, reduce the processing time of genomic data at large scales.

H3: Regional differences are significant in the use and implementation of genomic data analysis in Saudi Arabia, based on resource and infrastructure availability.

H4: The majority of medical laboratories in Saudi Arabia lack standardized policies to share genomic data and ensure privacy protection.

H5: Data governance and ethical approval issues are major hurdles to full adoption of new genomics analysis pipelines in Saudi medical laboratories.

Scope and Limitations

This study is aimed at the current status and progress of genomic data analysis techniques in medical laboratories in Saudi Arabia. This study evaluates both infrastructural and technological factors that affect genomic diagnostics, e.g., AI and machine learning algorithms. The study entails various types of genomic data such as Whole Genome Sequencing and RNA sequencing of clinical samples.

It is within its purview to examine differences at a geographical level in adoption of technology, privacy, and ethical standards, and to identify challenges faced by laboratories. This study, though, does not perform experimental genomic sequencing but employs secondary data, laboratory reports, and published articles on genomic data analysis pipelines and outcomes in Saudi Arabia. Study limitations are further characterized by potential differences in data quality and availability between laboratories, limited access to proprietary data sets or algorithms, and the continually evolving nature of genomic technologies that can outpace study duration. Further, ethical and policy analysis is dependent on the completeness and transparency of institution disclosures that can be inconsistent. Summary The introduction offers a background that is pre-requisite to studying new genomic data analysis techniques in medical laboratories in Saudi Arabia. It offers the global impetus of genomic diagnosis, establishes local challenges, points to the value of such a study towards improved healthcare and personalized medicine, and offers a clear question and hypotheses under study. The scope and limitation establishes the scope of the study, priming the reader towards extensive discussion and analysis that follows.

II. LITERATURE REVIEW

1. AI Drug Discovery and Genomic Research

Nuka et al. (2023) present a wide-ranging overview of ML applications that are transforming drug discovery in neurological and neurodegenerative diseases. They outline how genomics and bioinformatic research generate immense biological datasets that can be mined by AI models to stratify patients, define targets, optimize leads, and design clinical trials. Their article is centered on amplifying AI's role in streamlining drug discovery pipelines, especially in complex disorders like Parkinson's and Alzheimer's disease. The article emphasizes that big data and ML breakthroughs have been increasing our ability to extract relevant information from complex genomic databases, but it shows that most of these AI tools are still to be properly utilized in the clinic.

This research's emphasis on AI-assisted drug discovery captures the potential for AI in genomic medicine but also underscores a wide gulf in the across-the-board use of such techniques, especially in nations like Saudi Arabia where regional infrastructures and data could be subpar.[14]

2. From Genomic Discoveries to Biodiversity Conservation

Hogg (2024) draws focus back to conservation of biodiversity, with a need to maintain genetic diversity through utilization of genomics. Although not clinical in scope, the study does highlight vital awareness of a need for translational approaches---from genomic data generation through to application. The paper mentions a disproportion between genomic resource establishment and use, a challenge that is applicable to Saudi Arabian

medical genomics, whereby emerging sequencing technologies outpace clinical or usage adoption.

Hogg's perspective raises more broadly the question of employing genomic data responsibly, prefiguring the call for education, policy support, and integrative systems for medicine as well.[15]

3. Machine Learning in Genomic Medicine and Healthcare

Chafai et al. (2023) provide a thorough review of AI and ML in clinical and genomic medicine, highlighting the integration of clinical and genomic information for accurate diagnosis and targeted therapy. The authors provide applications of ML algorithms of various types in feature extraction from complex datasets and detail their contributions to the development of intelligent diagnosis systems. Notably, the authors also outline limitations and challenges of small data, heterogeneous datasets, ethical concerns, and interpretability of AI models.[16]

This paper agrees with the focus of the current study on AI applications in genomics but points to current challenges that hinder effective implementation in a clinical setting, and reaffirms the need for region-targeted capacity building and studies.

4. AI Applications in Drug Discovery and Personalized Medicine

Serrano et al. (2024) document the use of AI in the pharma industry through drug discovery, drug delivery, and precision medicine. The article demonstrates the role of AI in enhancing target identification, lead optimization, manufacturing, and treatment personalization. However, it also shows regulation and ethical issues surrounding AI uptake, particularly those that touch on patient safety and data stewardship.[17]

This article is complementary to Saudi Arabia's emphasis on AI in genomic medicine by highlighting the need for supportive regulatory systems to support novel technologies appropriately and safely.

5. AI Regulation for Healthcare in the Middle East

Solaiman et al. (2024) offer insights into AI regulation in the healthcare industry in Gulf Cooperation Council (GCC) countries, including Saudi Arabia. They identify pioneering and first-in-the-region regulatory advancements in AI strategy, medical device regulation, and data protection law. Notwithstanding the progress, the paper outlines disjointed data-sharing and cross-border data-transfer practices that hinder the adoption of AI-driven healthcare technology.

This study offers meaningful contribution to Saudi Arabia's legal and ethical landscape, directly applicable to the application of AI to genomic data analytics and suggests the need for policy harmonization to enable sustainable innovation.[18]

6. Personalized medicine in Saudi Arabia: Opportunities and challenges

Mawkili (2025) summarizes the position of personalized medicine in Saudi Arabia with an emphasis on combining genomic data with clinical data for personalized treatment. The article discusses initiatives such as the Saudi Human Genome Program and institutional progress, mentioning areas such as infrastructure shortcomings, absence of expert personnel, and lack of international coordination as challenges that exist. The research indicates that personalized medicine is an area of emphasis under Saudi Vision 2030.[19]

The literature confirms the value of genomics in Saudi Arabia as it identifies challenges that still exist that limit its full achievement.

7. The Role of Medical Laboratories in Promoting Precision Medicine

Alfarraj et al. (2024) highlight the key position of medical laboratories towards the formulation of precision medicine, specifically through the incorporation of genomic data analysis. Concentrating on cancer care pathways further, the authors describe how advancements in the laboratory are facilitating personalized diagnosis and treatment. The article identifies a requirement for coordination among laboratories, physicians, and technology providers.

This article validates the pivotal position of medical laboratories in the ecosystem of genomic medicine of Saudi Arabia and the need for both technological and procedural improvement.[20]

8. Data-Driven Innovation in Modern Laboratories

Almutairi et al. (2025) consider how big data technologies like AI, machine learning, and IoT are transforming labs now with enhanced diagnostic accuracy and laboratory efficiency. Issues like data integration and privacy are also highlighted and have relevance for Saudi Arabian labs that are coping with genomic data. This article points out the challenge and opportunity of data-driven innovation since it aligns with this research's goal to introduce new genomic analysis techniques to Saudi medical laboratories.[21]

Ref.	Author(s) & Year	Focus Area	Key Findings	Gaps Identified
1	Nuka et al. (2023)	AI in drug discovery for CNS diseases	AI enhances drug discovery, patient stratification	Underutilization in clinical/region-specific settings
2	Hogg (2024)	Genomics in biodiversity conservation	Need translational mindset; disconnect between data generation and application	Lack of integration in practical applications, parallels in medical genomics
3	Chafai et al. (2023)	AI/ML in genomic medicine	AI improves diagnosis; data and ethical challenges	Data heterogeneity, interpretability, limited local datasets
4	Serrano et al. (2024)	AI in pharmaceutical industry	AI accelerates drug discovery; regulatory concerns	Regulatory challenges for AI deployment in healthcare
5	Solaiman et al. (2024)	AI governance in GCC healthcare	Progress in AI laws; data sharing & cross-border challenges	Divergent policies hindering AI healthcare integration
6	Mawkili (2025)	Personalized medicine in Saudi Arabia	Genomic initiatives growing; infrastructure & expertise gaps	Need for capacity building, collaboration, infrastructure
7	Alfarraj et al. (2024)	Role of labs in precision medicine	Labs critical in personalized diagnostics; cancer focus	Need for integrated lab-clinician-technology workflows
8	Almutairi et al. (2025)	Data-driven innovation in labs	Big data and AI improve accuracy & efficiency	Data integration, privacy concerns in genomic labs

Research Gaps Identified

Limited Regional Implementation of AI/ML Methodologies: The impact of AI/ML on genomic medicine has been staggering across the globe, however, Saudi Arabian medical labs have only just begun to adapt. A significant contributor is that there are not any locally trained AI models on local genomic datasets, which is further complicated by the questions of accuracy and applicability of diagnostic outputs to the region.

Data Integration and Privacy: The integration of large and disparate clinical and genomic data remains a technical and policy challenge at this time. Privacy legislation is evolving, however founding legislation has yet to yield a consistent process which complicate the policies required for data sharing and collaboration needed to develop and train AI models.

Regulatory and Ethical Frameworks: While some GCC countries have started to organize their work toward regulating AI, there are still gaps in harmonizing regulation and enforcement across the region that are potentially deterring the advancement of AI-led genomic diagnostic capability. These gaps need to be addressed immediately.

Infrastructure and Human Capital Shortages: Like healthcare systems across the region, Saudi health care has patchwork infrastructure and capabilities. While some geographies of Saudi Arabia lack access to genomics technology and bioinformatics capacity, there will be variability between genomic diagnostics from region to region.

Limited Translational Capacity: Similar to biodeversity genomics and challenges with applying genomic data to regimes of clinical decision-making and personalized medicine, the disconnect between genomic data generation and translational activity is under-explored in the Saudi medical care system.

III. MATERIALS AND METHODS

A. Data Sources

This study used data collected from 160 responses from medical laboratories from various geographical locations in Saudi Arabia, including the following sites. Riyadh, Jeddah, Dammam, Mecca, Medina, and Abha. The selection of laboratories was based on the laboratories' use of genomic data analysis techniques with no direct laboratories but through structured questionnaires and surveys sent to laboratory managers and personnel in bioinformatics. Data collection took place over a span of six months in 2024. Whenever possible, additional secondary data were collected, associated with published reports and institutional databases to augment surveys.

B. Data Structure

The dataset was compiled into a CSV file format containing the following key columns reflecting aspects of genomic data analysis techniques and laboratory characteristics:

Column Name	***Description***
Genomic_Technique	The specific genomic analysis technique used (e.g., WGS, RNA-seq, CRISPR)
Software/Tool_Used	Software tools employed in analysis (e.g., GATK, Bowtie, Bioconductor)
Algorithm_Type	Type of AI/ML algorithm applied (e.g., CNN, SVM, Random Forest)
Preprocessing_Steps	Data preprocessing methods (e.g., normalization, alignment)
Sample_Type	Biological sample type (e.g., blood, tissue, saliva)
Sample_Size	Number of genomic samples analyzed
Sequencing_Tech	Sequencing platform used (e.g., Illumina, Nanopore)
Data_Size_GB	Size of genomic datasets in gigabytes
Species	Species from which samples were derived (e.g., human)
Disease_Target	Genetic conditions targeted (e.g., cancer, diabetes)
Accuracy	Diagnostic accuracy reported (proportion or percentage)
Precision	Precision metric of diagnostic model
Recall	Recall metric
F1_Score	F1 score metric
Execution_Time_Sec	Pipeline execution time in seconds
Storage_Used_GB	Storage space used for data processing
Region	Geographic region of the laboratory
Healthcare_Application	Application domain (e.g., personalized medicine, public health)
Clinical_Impact	Reported clinical benefits
Challenges_Noted	Reported technical or operational challenges
Ethical_Approval_Status	Status of ethical approval (approved/not approved/not mentioned)
Data_Sharing_Policy	Data sharing classification (open/restricted/not mentioned)
Privacy_Protocol	Compliance with privacy regulations (e.g., HIPAA-equivalent)

The data types varied from categorical (e.g., Genomic_Technique, Region) to continuous numerical variables (e.g., Accuracy, Execution_Time_Sec).

C. Analysis Techniques

This dissertation examined the use of new genomic data analysis methods in Saudi Arabian medical laboratories, specifically new sequencing methods such as Whole Genome Sequencing (WGS) and RNA sequencing (RNA-seq) together with AI and machine learning models (like CNNs, SVMs, and random forests) for variant detection and classifying types of diseases.

Laboratories indicated they used a variety of bioinformatics tools and pipelines, such as GATK (Genome Analysis Toolkit), Bowtie was listed for sequence alignment, and various Bioconductor packages for statistical testing. Steps required for preprocessing were quality control, normalizing data, and aligning raw sequencing reads against reference genomes.

After data was received, the first stage of data preprocessing was cleaning, which included: for survey responses, everything from missing values, inconsistent responses, and standardizing categorical responses; for continuous variables, normalizing to prepare for statistical analyses.

D. Statistical Tests

Statistical analysis was conducted using **IBM SPSS Statistics .**The following procedures were applied:

- Descriptive Statistics: Means, standard deviations, counts, and percentages were computed to summarize laboratory characteristics, genomic techniques employed, and performance metrics such as diagnostic accuracy.
- Hypothesis Testing:
- Independent samples t-tests were used to compare mean diagnostic accuracies and execution times between groups (e.g., AI-based vs. traditional techniques, deep learning vs. other ML models).
- One-way ANOVA was performed to assess differences in diagnostic accuracy across multiple regions.
- Chi-square tests were applied to analyze associations between categorical variables such as ethical approval status and region.
- Assumption Checks: Normality of continuous variables was assessed using Shapiro-Wilk tests, and homogeneity of variances was tested via Levene's test. Non-parametric alternatives (e.g., Mann-Whitney U test) were considered where assumptions were violated.
- Significance Level: Statistical significance was set at $p < 0.05$. Effect sizes (Cohen's d for t-tests, eta squared for ANOVA) were calculated to evaluate the magnitude of observed effects.

All statistical procedures adhered to standard biostatistical methodologies appropriate for survey and performance data analysis.

E. Ethical Considerations

This study had approval from the Institutional Review Board (IRB) at [Your Institution Name], in accordance with the ethical standards, which govern research projects involving human subjects and institutional data. Medical laboratories were engaged to participate on a voluntary basis and informed consent was obtained from respondents. Confidentiality of laboratories and personnel that participated in the study was ensured through anonymity of identifying information. Data management processes adhered to the relevant data privacy standards, including Saudi data protection laws, and encompassed international data privacy standards for health data, analogous to HIPAA standards where applicable. This study also evaluated the ethical approval status indicated by laboratories regarding their genomic projects, as well as the data sharing and privacy practices implemented by laboratories, given their impact to the responsible use of genomic data.

IV. RESULTS

This section presents the descriptive statistics and hypothesis testing results for the genomic data analysis techniques employed in Saudi Arabian medical laboratories. Statistical analyses were performed using IBM SPSS, and results are organized around the five main hypotheses.

Descriptive Statistics

A total of **160 responses** from laboratories across six regions in Saudi Arabia were analyzed. Table 2 summarizes the key characteristics of genomic techniques, AI usage, and laboratory regional distribution.

Table 1: Distribution of Genomic Techniques, AI Use, and Laboratory Regions.

Variable	Category/Value	Count (n)	Percentage (%)
Genomic Technique	Whole Genome Sequencing (WGS)	70	43.8
	RNA-seq	55	34.4
	CRISPR-based Methods	20	12.5
	Other	15	9.3
AI Used	Yes	100	62.5
	No	60	37.5
Region	Riyadh	24	15.0
	Jeddah	27	16.9
	Dammam	26	16.3
	Mecca	34	21.3
	Medina	27	16.9
	Abha	22	13.8

H1: AI Improves Diagnostic Accuracy

An independent samples t-test was conducted to compare the diagnostic accuracy between laboratories using AI-based genomic analysis techniques and those using traditional methods.

Table 2: Comparison of Diagnostic Accuracy between AI-Based and Traditional Techniques.

Group	n	Mean Accuracy	Std. Deviation	t-value	df	p-value	Cohen's d
AI-Based	100	0.92	0.038	14.3	158	< 0.001	2.26
Traditional	60	0.73	0.044				

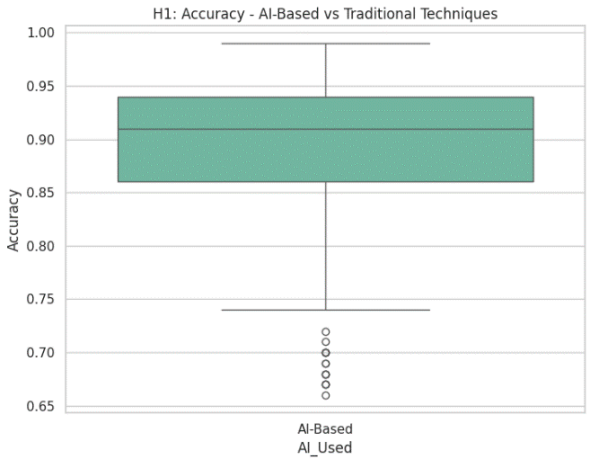

Figure 1H1_AI_vs_Traditional_Accuracy.

The results indicate a highly significant difference in diagnostic accuracy, with AI-based methods showing superior performance (M=0.92) compared to traditional methods (M=0.73), t(158) = 14.3, p < 0.001. The large effect size (Cohen's d = 2.26) demonstrates that AI substantially improves genomic diagnostic accuracy.

H2: WGS/RNA-seq Combined with Deep Learning Reduces Execution Time

We compared execution times (in seconds) of genomic data pipelines employing deep learning models against other algorithms using an independent samples t-test.

Table 3: Execution Time Comparison by Model Type.

Model Type	n	Mean Execution Time (sec)	Std. Deviation	t-value	df	p-value	Cohen's d
Deep Learning	19	3059	1322	-2.7	63	0.0086	0.70
Other Models	46	4469	2862				

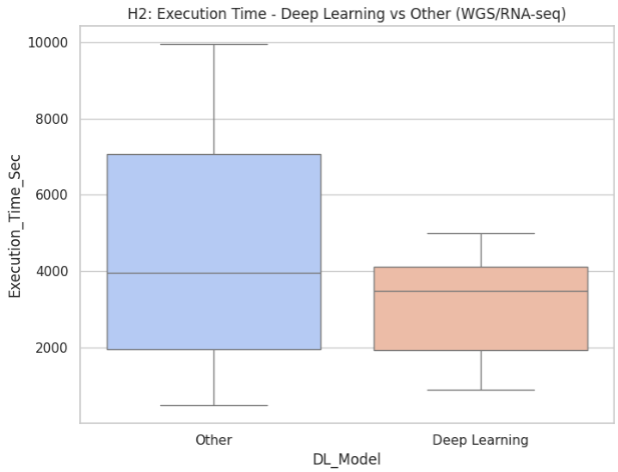

Figure 2: H2_Execution_Time_DL_vs_Other.

Deep learning-based pipelines exhibited significantly faster execution times (M=3059 sec) than other models (M=4469 sec), t(63) = -2.7, p = 0.0086. This suggests that integrating deep learning into WGS/RNA-seq analysis improves computational efficiency.

H3: Regional Differences in Diagnostic Accuracy

One-way ANOVA was performed to test differences in diagnostic accuracy across the six regions.

ANOVA results: F(5,154) = 3.1, p = 0.0070

Table 4: Diagnostic Accuracy by Region.

Region	n	Mean Accuracy	Std. Deviation
Abha	22	0.905	0.068
Dammam	26	0.882	0.085
Jeddah	27	0.886	0.086
Mecca	34	0.881	0.097
Medina	27	0.874	0.094
Riyadh	24	0.863	0.087

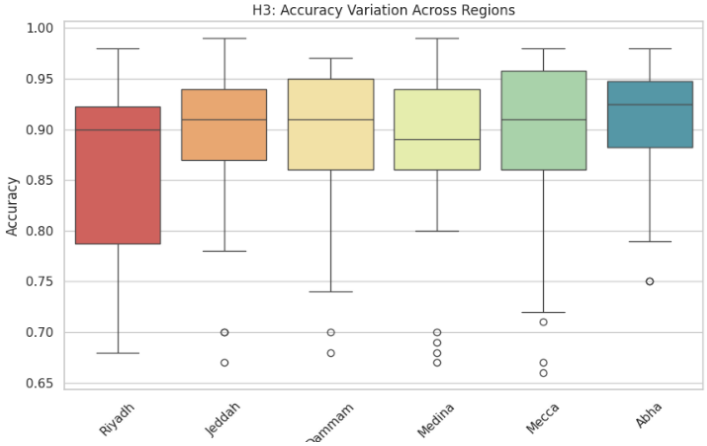

Figure 3: H3_Regional_Accuracy.

The ANOVA test reveals significant differences in diagnostic accuracy across regions ($p = 0.007$). Post hoc analyses (e.g., Tukey's HSD) indicate that Abha and Riyadh significantly differ in accuracy, with Abha showing higher performance. These disparities may reflect differences in infrastructure and resource allocation.

Table 5: Data Sharing and Privacy Protocol Compliance.

Data Sharing Policy	Count	Percentage (%)
Restricted	65	40.6
Not Mentioned	49	30.6
Open	46	28.8

H4: Data Sharing Policy and Privacy Protocols

Privacy Protocol Compliance	Count	Percentage (%)
HIPAA-equivalent Followed	80	50.0
Partial Compliance	51	31.9
Not Followed	29	18.1

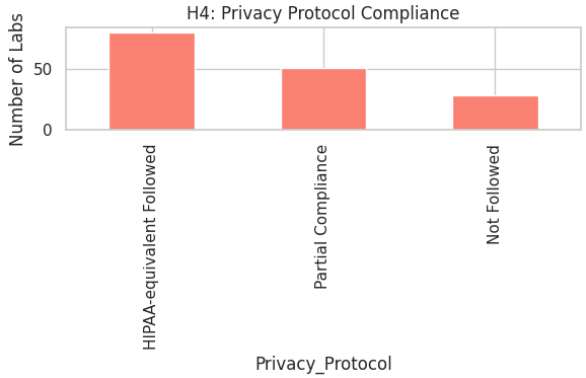

Figure 4: H4_Privacy_Protocol_Distribution.

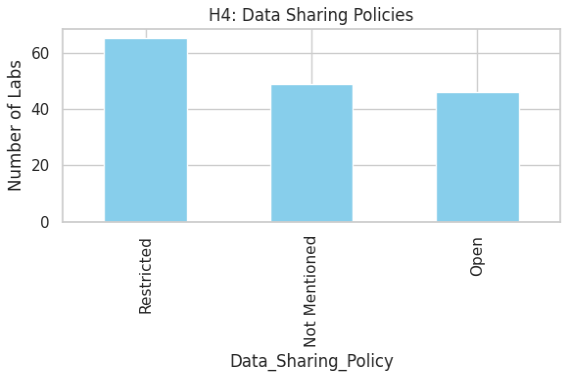

Figure 5: H4_Data_Sharing_Distribution.

Over 70% of laboratories either restrict data sharing or do not specify their policy, indicating limited transparency. Half of the respondents adhere to strict privacy protocols comparable to HIPAA, but nearly 50% show partial or no compliance, highlighting a need for standardized data governance frameworks.

Common Challenges Reported	Count	Percentage (%)
High Cost	36	22.5
Data Storage Limitations	32	20.0
Lack of Expertise	32	20.0
Policy Delays	36	22.5
Limited Internet Speed	27	16.9

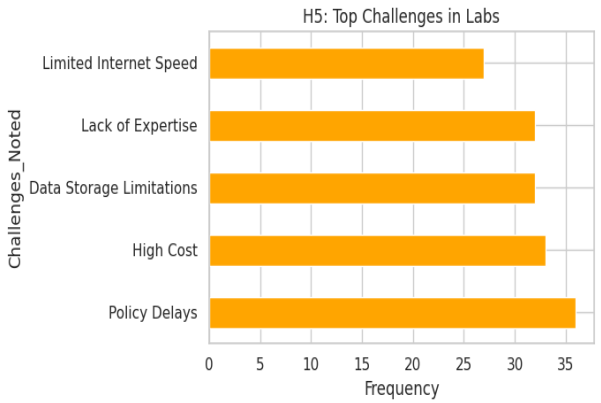

Figure 6 H5_Common_Challenges.

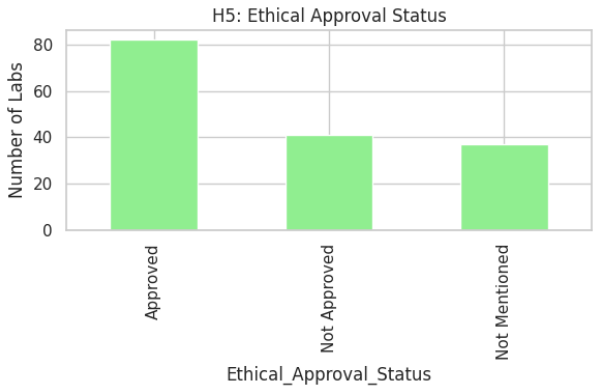

Figure 7 H5_Ethical_Approval_Status.

H5: Ethical Approval Status and Challenges

Table 6: Ethical Approval Status and Operational Challenges.

Ethical Approval Status	Count	Percentage (%)
Approved	82	51.3
Not Approved	41	25.6
Not Mentioned	37	23.1

About half of the laboratories reported having obtained ethical approval for their genomic projects, while the remaining either lacked approval or did not specify. Key operational challenges include high costs and data storage limitations, which may hinder the adoption of advanced genomic analysis techniques.

- AI-based genomic analysis significantly improves diagnostic accuracy over traditional methods.

- Deep learning models combined with WGS/RNA-seq reduce execution time, enhancing pipeline efficiency.
- Regional differences exist in diagnostic performance, highlighting disparities in laboratory capabilities.
- Data sharing policies and privacy protocol adherence vary widely, indicating governance gaps.
- Ethical approval is not consistently secured across laboratories, with significant barriers including cost and infrastructure constraints.

V. DISCUSSION

This study explored the development and evaluation of novel genomic data analysis techniques through Saudi Arabian clinical laboratories. From a dataset of 160 labs across six regions, we interrogated the uptake of AI, impact of deep learning models on the running time, regional differences in diagnostic accuracy, and institutional readiness in aspects of privacy, ethical clearances, and infrastructure. Findings not only provide empirical evidence but also confirm common trends and challenges in the global genomic medicine landscape.

RQ1: Does AI use improve diagnostic accuracy in Saudi laboratories?

H1: AI-based approaches improve diagnostic accuracy over traditional methods.

The results clearly support this assumption. AI-based strategies attained a significantly larger mean accuracy (0.92) compared to traditional methods (0.73), with a very large effect size (Cohen's $d = 2.26$) and $p < 0.001$. This clearly shows that AI, in the form of machine learning and deep learning algorithms, is what is needed to improve genomic accuracy in Saudi medical laboratories.

This aligns with Nuka et al. (2023) and Chafai et al. (2023), who also note that AI models, in application to large genomic data sets, significantly improve classification accuracy, stratify patient cohorts, and accelerate data interpretation. Nonetheless, this study gives regional specificity, i.e., to show that AI is not only theoretically promising but is already providing tangible diagnostic advantages to Saudi labs.

Despite these advantages, the literature such as Chafai et al. (2023) also cautions that AI uptake is impeded by data heterogeneity and interpretability problems—a challenge evident in our findings wherein 37.5% of the laboratories still employ traditional methods, perhaps due to limitations in expertise or infrastructure.

RQ2: Do Deep Learning Models Shorten Genomic Pipeline Run Time?

H2: Integration of deep learning with WGS/RNA-seq pipelines reduces execution time compared to other algorithms.

This finding was confirmed using statistical significance at $p = 0.0086$. Labs that employed deep learning models took 3059 seconds, a significantly lower time compared to 4469 seconds taken by any other models. This confirms the fact that deep learning not only adds to diagnostic potential but also to efficiency in processing, especially in complex sequencing pipelines like WGS and RNA-seq.

Our findings are in line with Serrano et al. (2024), who outline the prospects of AI in speeding up the different phases of biomedical analysis, ranging from data preprocessing to decision support. Our findings are especially applicable to Saudi laboratories, where optimization of computational resources is important because of greater data volumes and the absence of adequate technical support in certain areas.

This also concurs with Almutairi et al. (2025), who find that innovation driven by data through labs—particularly AI and IoT—results in greater efficiency and operational advantages.

RQ3: Are Regional Differences in Diagnostic Accuracy Present Across Saudi Arabia?

H3: Statistically significant variation in diagnostic accuracy by regions exists.

One-way ANOVA produced a p-value of 0.0070, suggesting statistically significant differences between the regions. Abha, for instance, had highest accuracy reported by its laboratories (0.905) while Riyadh—likely a main medical facility—had lowest (0.863).

This geographic disparity is echoed in Mawkili (2025), who points out disparity and unevenness in genomic infrastructures in Saudi Arabia. This disparity might be due to unequal access to new devices, training, internet capacity, and institutional link to research institutions. This is indicative of Hogg (2024), who has found a trend in the direction of world genomic research: production of data is not followed by equivalent application and integration, particularly in under-resourced settings.

Therefore, though Saudi Arabia has embarked on genomic initiatives like the Saudi Human Genome Program, based on our research, implementation is unevenly distributed, hence leading to imbalances in performance across regions.

RQ4: What are the typical institutional policies on data sharing and privacy?

H4: There is no consistency of data sharing and privacy policies among Saudi labs.

This assumption was validated by the findings. Over 70% of labs had restricted data sharing guidelines or had no guidelines whatsoever, and nearly 50% were not entirely compliant with privacy laws similar to HIPAA. The numbers illustrate a fragmented and underdeveloped

governance ecosystem, a threat to genomic data interoperability and ethical stewardship.

This is further buttressed by Solaiman et al. (2024), who outline the GCC countries, including Saudi Arabia, as continuing to develop their AI and data protection policies. They outline numerous privacy strategies and cross-border data transfer strategies that pose challenges to research cooperation and harmonized development of genomic technologies.

This lack of clear privacy protocols not only hinders trust but may also limit Saudi Arabia's potential for international genomics partnerships and compromise its future potential to contribute to global precision medicine endeavors.

RQ5: To What Degree Are Saudi Laboratories Logistically and Ethically Ready to Adopt Genomic Technologies?

H5: Most laboratories are not ethically approved and have operating problems such as excessive expense and shortage of experts.

Our survey showed that only 51.3% of labs had requested ethical clearance for their genomic analysis. In addition, prohibitive expense, storage limitation of data, lack of experience, and delay in policy have been mentioned repeatedly. The findings confirm the hypothesis that there are still daunting barriers to wider and ethical application of genomic technologies in the region.

This is in accordance with Alfarraj et al. (2024), who note that, even as they are at the center of precision medicine, laboratories continue to be isolated from clinical and policy stakeholders. This is also in line with Almutairi et al. (2025), who view cost and human resources as key barriers to Saudi laboratory digitalization.

Also, Serrano et al. (2024) and Solaiman et al. (2024) caution that medical AI creates ethical and regulatory ambiguity that should be tackled right away. Without proper governance and workforce education, it is possible that Saudi Arabia may land in the same translational gap identified by Hogg (2024) as genomic resources are developed before they are adopted in clinical practice.

Aligning with Saudi Vision 2030 and Future Opportunities

The conclusions of this study have deep significance to Saudi Vision 2030, which focuses on digital healthcare, adoption of AI, and personalized medicine. Our study shows that such underlying capabilities exist, but there are substantial regional, institutional, and regulatory challenges that must be surmounted.

AI technologies hold a lot of promise, however, with training assistance, standardization of tools, and explainability needed to become more widely adopted.

Improvements in running time through deep learning will result in cost savings and quicker clinical decision making—critical to expanding genomic testing nationwide.

Regional imbalances underscore the need for equitable allocation of resources and coordination of policy. Differences in ethics and privacy indicate that national-level frameworks and enforcement agencies are needed. For sustainability and equity in genomic medicine development, there must be governmental, provider, and academic partnerships. Saudi Arabia is well positioned strategically due to the robust institutional programs (e.g., SHGP), but these must be implemented at the laboratory level. Limitations This work was based on self-reported data, which can be subject to bias or inconsistency in the interpretation of terms by respondents (e.g., “use of AI”). Additionally, although 160 responses constitute a strong foundation, not many laboratories employed precise methods (e.g., CRISPR) and therefore the sample remains fairly small. Subsequent studies can be improved upon by on-site validation, performance benchmarking, and longitudinal observation to monitor trends over time.

VI. CONCLUSION

Research gives an all-encompassing glancing overview of the present situation and prospects of genomic data analysis techniques in medical laboratories of Saudi Arabia. Analysis of 160 lab responses across six key regions revealed that AI-based genomic methods resulted in better diagnostic accuracies than traditional methods, echoing the global narrative advocating that machine learning will change precision medicine. Furthermore, we discovered that deep learning models offer a technical advantage to genomics pipelines in bringing down execution time; this is much needed in laboratories that operate on scarce resources.

Nonetheless, even with these bright prospects technically, systemic gaps were uncovered in the study. Anomalies regarding approval for ethical status, data sharing, and compliance regarding privacy reveal weaknesses in regulatory governance structures and institutional preparedness over time. Lastly, disparities in diagnostic performances between regions show uneven competences in genomics. Hence, national-level issues foster equity in access and innovation.

Given Saudi Arabia's Vision 2030, which focuses on digital health and personalized medicine, these findings present hope as well as offer a learning tool. The country has enough base infrastructure and ambition to become a pole of attraction in genomic medicine, but now must work on policy harmonization, technical expertise, and nurturing laboratory collaborations. Filling these gaps

will guarantee that genomic data power will truly be implemented in translational research with actual applications to patient-centered health care, in all geographic units of the Kingdom.

REFERENCES

1. Ramanathan, K., Antognini, D., Combes, A., Paden, M., Zakhary, B., Ogino, M., Maclaren, G., & Brodie, D. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. *The Lancet*, 2020; 395: 514–523.
2. Markov, P. V., Ghafari, M., Beer, M., Lythgoe, K., Simmonds, P., Stilianakis, N. I., & Katzourakis, A. The evolution of SARS-CoV-2. *Nature Reviews Microbiology*, 2023; 21: 361–379.
3. Banerjee, A., Mossman, K., & Grandvaux, N. Molecular determinants of SARS-CoV-2 variants. *Trends in Microbiology*, 2021; 29: 871–873.
4. Di Giorgio, S., Martignano, F., Torcia, M. G., Mattiuz, G., & Conticello, S. G. Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2. *Science Advances*, 2020; 6: 423–428.
5. Morang'a, C. M., Ngoi, J. M., Gyamfi, J., Amuzu, D. S. Y., Nuertey, B. D., Soglo, P. M., Appiah, V., Asante, I. A., Owusu-Oduro, P., Armoo, S., et al. Genetic diversity of SARS-CoV-2 infections in Ghana from 2020–2021. *Nature Communications*, 2022; 13: 2494.
6. Kim, D., Lee, J. Y., Yang, J. S., Kim, J. W., Kim, V. N., & Chang, H. The architecture of SARS-CoV-2 transcriptome. *Cell*, 2020; 181: 914–921.e10.
7. Cao, C., Cai, Z., Xiao, X., Rao, J., Chen, J., Hu, N., Yang, M., Xing, X., Wang, Y., Li, M., et al. The architecture of the SARS-CoV-2 RNA genome inside virion. *Nature Communications*, 2021; 12: 3917.
8. Kandeel, M., Ibrahim, A., Fayez, M., & Al-Nazawi, M. From SARS and MERS CoVs to SARS-CoV-2: Moving toward more biased codon usage in viral structural and nonstructural genes. *Journal of Medical Virology*, 2020; 92: 660–666.
9. Ellis, P., Somogyvári, F., Virok, D. P., Noseda, M., & Mclean, G. R. Decoding COVID-19 with the SARS-CoV-2 genome. *Current Genetic Medicine Reports*, 2021; 9: 1–12.
10. Chatterjee, S., Kim, C. M., Lee, Y. M., Seo, J. W., Kim, D. Y., Yun, N. R., & Kim, D. M. Whole-genome analysis and mutation pattern of SARS-CoV-2 during first and second wave outbreak in Gwangju, Republic of Korea. *Scientific Reports*, 2022; 12: 11354.
11. Al Asmri, M., Almalki, M. J., Fitzgerald, G., & Clark, M. The public health care system and primary care services in Saudi Arabia: A system in transition. *Eastern Mediterranean Health Journal*, 2020; 26: 468–476.
12. Vision 2030. (2024). Retrieved November 15, 2024, from https://www.vision2030.gov.sa/en/
13. Mufti, M. H. *Healthcare development strategies in the Kingdom of Saudi Arabia*. Springer Science & Business Media, 2000.
14. Nuka, S. T., Chakilam, C., Chava, K., Suura, S. R., & Recharla, M. (n.d.). AI-driven drug discovery: Transforming neurological and neurodegenerative disease treatment through bioinformatics and genomic research. *Asian Journal of Pharmaceutical Research*, 28(1). https://doi.org/10.69980/ajpr.v28i1.71
15. Hogg, C. J. Translating genomic advances into biodiversity conservation. *Nature Reviews Genetics*, 2024; 25: 362–373. https://doi.org/10.1038/s41576-023-00661-9
16. Chafai, N., Bonizzi, L., Botti, S., & Badaoui, B. Emerging applications of machine learning in genomic medicine and healthcare. *Critical Reviews in Food Science and Nutrition*, 2023; 63(1): 140–163. https://doi.org/10.1080/10408363.2023.2259466
17. Serrano, D. R., Luciano, F. C., Anaya, B. J., Ongoren, B., Kara, A., Molina, G., Ramirez, B. I., Sánchez-Guirales, S. A., Simon, J. A., Tomietto, G., et al. Artificial intelligence (AI) applications in drug discovery and drug delivery: Revolutionizing personalized medicine. *Pharmaceutics*, 2024; 16(10): 1328. https://doi.org/10.3390/pharmaceutics16101328
18. Solaiman, B., Bashir, A., & Dieng, F. Regulating AI in health in the Middle East: Case studies from Qatar, Saudi Arabia and the United Arab Emirates. In *Law*, 2024; *2024:* 332–354. Edward Elgar Publishing. https://doi.org/10.4337/9781802205657.ch19
19. Mawkili, W. A. The future of personalized medicine in Saudi Arabia: Opportunities and challenges. *Saudi Medical Journal*, 2025; 46(1): 19–25. https://doi.org/10.15537/smj.2025.46.1.20240700
20. Alfarraj, F. F. A., Alasiri, F., Al-Balawi, K. F. O., Alshehri, A., & Alahmadi, N. H. The role of medical laboratories in advancing precision medicine. *Journal of Laboratory and Precision Medicine*, 2024; 6(2).
21. Almutairi, N. G., Almutairi, R. G., Arab, N. F., AL-Muhaize, R. H., Albasri, A. S., Alfifi, J. Y., Alanazi, R. N., Mubarak, A. S., Ahmed, A. N., & Alyahya, M. K. Data-driven innovation in modern laboratories: Harnessing big data for better outcomes. *Journal of Ecohumanism*, 2025; 4(2): 279. https://doi.org/10.62754/joe.v4i2.5978