International Journal of
Computer Science and Artificial Intelligence
https://www.ioro.org/ijcsa

AI-Driven Universal DNA Extraction and Quality Prediction Framework for Cross-Kingdom Genomic Applications

Mulugeta Tilahun Bekele

Department of Information Technology, College of Informatics, University of Gondar
mulugeta.tilahun@gmail.com mulugeta.tilahun@uog.edu.et

Abstract:
High-quality genomic DNA is fundamental for molecular diagnostics, precision medicine, agricultural biotechnology, microbial genomics, and biodiversity conservation. However, existing DNA extraction protocols are generally optimized for specific organisms, resulting in inconsistent DNA yield, purity, processing time, and cost when applied across different biological kingdoms. Furthermore, conventional laboratory methods lack intelligent mechanisms for predicting DNA quality before downstream genomic analyses.
This study aimed to develop and evaluate an AI-Driven Universal DNA Extraction and Quality Prediction Framework capable of extracting high-quality genomic DNA from bacterial, plant, and mammalian samples while accurately predicting DNA quality using artificial intelligence.
A mixed-methods research design was employed using representative bacterial (Escherichia coli), plant (Arabidopsis thaliana), and mammalian blood samples. An eco-friendly universal DNA extraction protocol was integrated with supervised machine learning algorithms. The proposed framework was compared with CTAB, phenol–chloroform, and commercial silica column methods using DNA yield (ng/µL), purity (A260/A280 and A260/A230), DNA integrity, extraction time, reagent cost, reproducibility, PCR amplification success, and AI prediction accuracy.
The proposed framework achieved an average DNA purity of 1.87–1.95 (A260/A280), increased DNA yield by 22.8%, reduced extraction time by 34.5%, lowered reagent cost by 41.2%, and achieved a 97.1% PCR amplification success rate across bacterial, plant, and mammalian samples. The AI-based quality prediction model attained 96.8% prediction accuracy, providing reliable and interpretable assessments of DNA quality for genomic applications.
The proposed AI-driven universal framework provides a scalable, environmentally sustainable, and cost-effective solution for cross-kingdom genomic DNA extraction and intelligent quality prediction. Its superior performance across multiple biological sample types demonstrates its potential to standardize molecular biology workflows and support clinical genomics, agricultural biotechnology, environmental DNA research, and biodiversity conservation.

Keywords: Artificial Intelligence, Universal DNA Extraction, DNA Quality Prediction, Cross-Kingdom Genomics, Machine Learning.

1. Introduction

1.1 Background of the Study

Genomic DNA extraction is the foundational step in molecular biology, genomics, forensic science, clinical diagnostics, agricultural biotechnology, environmental monitoring, and biodiversity conservation. The quality of extracted DNA directly influences the reliability of downstream applications such as polymerase chain reaction (PCR), quantitative PCR (qPCR), next-generation sequencing (NGS), whole-genome sequencing,

metagenomics, gene expression analysis, molecular diagnostics, DNA barcoding, and genome editing. DNA with high purity, sufficient concentration, minimal fragmentation, and low levels of contaminants is essential for producing accurate and reproducible genomic results (Green & Sambrook, 2019).

Despite substantial advances in molecular biology, no single DNA extraction protocol consistently performs well across diverse biological kingdoms. Most conventional extraction methods have been optimized for specific sample types because microorganisms, plants, and animals differ considerably in cellular architecture, biochemical composition, and the presence of inhibitory compounds. Gram-positive bacteria possess thick peptidoglycan cell walls, Gram-negative bacteria contain outer membrane lipopolysaccharides, plant tissues contain cellulose, lignin, polysaccharides, polyphenols, and secondary metabolites, whereas mammalian blood contains proteins, lipids, and heme compounds that can inhibit enzymatic reactions. Consequently, extraction procedures that perform effectively for one biological group often yield poor DNA quality or reduced efficiency when applied to another (Tan & Yiap, 2009).

Currently, laboratories employ multiple extraction techniques, including cetyltrimethylammonium bromide (CTAB), phenol–chloroform extraction, silica spin-column purification, magnetic bead-based extraction, salting-out procedures, and commercial extraction kits. Each technique offers distinct advantages but also presents limitations regarding DNA yield, purity, extraction time, chemical toxicity, operational complexity, reagent cost, scalability, reproducibility, and environmental sustainability. Organic solvent-based methods frequently produce high DNA yields but involve hazardous chemicals such as phenol and chloroform, whereas commercial kits provide standardized procedures but significantly increase laboratory operating costs, particularly in resource-limited environments. CTAB-based protocols remain effective for plant tissues but often require protocol modifications when processing bacterial or animal samples, reducing their universal applicability (Sambrook & Russell, 2001; Green & Sambrook, 2019).

The increasing integration of genomics into precision medicine, agricultural improvement, infectious disease surveillance, wildlife conservation, and environmental DNA (eDNA) monitoring has created an urgent demand for a universal DNA extraction framework capable of processing diverse biological samples using standardized procedures. Such a framework should maintain high DNA quality while minimizing laboratory time, reducing operational costs, improving reproducibility, and decreasing the environmental impact associated with hazardous chemical waste. Developing environmentally sustainable ("green") extraction approaches is becoming increasingly important as research institutions adopt safer laboratory practices and sustainable biotechnology initiatives (UNESCO, 2021).

Artificial Intelligence (AI) has emerged as a transformative technology capable of optimizing complex laboratory workflows through machine learning, predictive analytics, and intelligent decision support. In molecular biology, AI algorithms have demonstrated remarkable capabilities in sequence analysis, protein structure prediction, variant detection, laboratory automation, genomic classification, and predictive modeling. However, relatively limited research has focused on applying AI to optimize DNA extraction protocols or predict DNA quality before downstream genomic analysis. Integrating AI into DNA extraction workflows enables the analysis of multidimensional laboratory variables—including sample type, lysis duration, reagent concentration, pH, temperature, incubation conditions, centrifugation speed, and purification parameters—to predict DNA quality with high accuracy and recommend optimal extraction conditions.

Unlike traditional laboratory optimization, which depends largely on trial-and-error experimentation, AI-driven predictive models can learn complex nonlinear relationships among extraction variables and quality outcomes. Supervised machine learning algorithms such as Random Forest, Extreme Gradient Boosting

(XGBoost), Support Vector Machine (SVM), Artificial Neural Networks (ANN), and ensemble learning methods have shown strong predictive performance for biological datasets because they effectively model variable interactions while maintaining robustness against noisy experimental data (Jordan & Mitchell, 2015). Incorporating explainable artificial intelligence (XAI) further enhances transparency by identifying the laboratory variables that most strongly influence DNA quality, thereby supporting evidence-based laboratory decision-making.An effective universal DNA extraction framework must therefore be evaluated using comprehensive and objective comparison parameters rather than relying solely on DNA concentration. Key performance indicators include DNA yield (ng/µL), purity measured by A260/A280 and A260/A230 absorbance ratios, DNA integrity assessed by agarose gel electrophoresis, PCR amplification success, sequencing compatibility, extraction time, reagent consumption, chemical toxicity, operational cost, reproducibility across laboratories, scalability for high-throughput processing, environmental sustainability, and AI prediction performance metrics such as accuracy, precision, recall, F1-score, mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2). Simultaneous evaluation of these biological, technical, economic, and computational parameters provides a more rigorous assessment of framework performance than conventional single-metric comparisons.

Therefore, this study proposes an AI-Driven Universal DNA Extraction and Quality Prediction Framework for Cross-Kingdom Genomic Applications that integrates an eco-friendly universal extraction protocol with explainable machine learning models capable of predicting DNA quality across bacterial, plant, and mammalian samples. The framework seeks to overcome the limitations of species-specific extraction methods by improving DNA yield, purity, integrity, reproducibility, cost efficiency, environmental sustainability, and intelligent laboratory decision-making. Ultimately, the proposed framework is expected to contribute toward standardized genomic workflows suitable for clinical diagnostics, agricultural biotechnology, microbial genomics, biodiversity research, environmental DNA monitoring, and precision genomic medicine.

1.2 Problem Statement

High-quality genomic DNA is a prerequisite for reliable molecular biology applications, including polymerase chain reaction (PCR), quantitative PCR (qPCR), next-generation sequencing (NGS), whole-genome sequencing, metagenomics, DNA barcoding, molecular diagnostics, precision medicine, agricultural genomics, and biodiversity conservation. The success of these downstream applications depends largely on the quality of extracted DNA, particularly its yield, purity, integrity, and freedom from contaminants. However, obtaining consistently high-quality DNA from diverse biological sources remains a major challenge because organisms from different kingdoms possess distinct cellular structures and biochemical compositions that require different extraction approaches (Green & Sambrook, 2019).

Current DNA extraction protocols—including cetyltrimethylammonium bromide (CTAB), phenol–chloroform extraction, silica spin-column purification, magnetic bead-based extraction, and commercial extraction kits—are primarily designed for specific sample types rather than universal application. Plant tissues contain cellulose, lignin, polysaccharides, and polyphenolic compounds that interfere with DNA purification, bacterial cells exhibit highly variable cell wall structures, and mammalian blood contains proteins, lipids, and heme-derived PCR inhibitors. Consequently, laboratories frequently employ multiple extraction protocols for different organisms, resulting in inconsistent DNA quality, increased laboratory complexity, higher reagent costs, longer processing times, reduced reproducibility, and greater environmental burdens due to hazardous chemical usage (Tan & Yiap, 2009).

Another significant limitation is that the performance of DNA extraction methods is often evaluated using only one or two indicators, such as DNA concentration or the A260/A280 absorbance ratio. Such limited assessment fails to provide a comprehensive understanding of extraction performance across biologically diverse samples. Critical comparison parameters—including DNA yield (ng/µL), purity (A260/A280 and A260/A230), DNA integrity, PCR amplification success, sequencing compatibility, extraction time, reagent consumption, operational cost, reproducibility, scalability, environmental sustainability, and robustness across bacterial, plant, and mammalian samples—are rarely evaluated simultaneously within a single standardized framework. As a result, there is insufficient evidence to determine which extraction strategy provides the best overall performance for cross-kingdom genomic applications.

Furthermore, laboratory optimization of DNA extraction remains largely empirical and depends on repeated trial-and-error experiments. Variables such as sample type, tissue mass, lysis duration, reagent concentration, pH, incubation temperature, centrifugation conditions, and purification steps interact in complex, nonlinear ways that are difficult to optimize manually. Existing protocols generally lack intelligent decision-support systems capable of predicting DNA quality before downstream analysis, increasing the likelihood of failed PCR amplification, poor sequencing performance, unnecessary reagent consumption, and repeated laboratory procedures. Although artificial intelligence (AI) has demonstrated remarkable success in genomics, bioinformatics, and biomedical prediction, its application to universal DNA extraction optimization and DNA quality prediction across multiple biological kingdoms remains limited (Jordan & Mitchell, 2015).

In addition, the growing demand for precision medicine, agricultural biotechnology, environmental DNA (eDNA) monitoring, microbial surveillance, and biodiversity genomics requires standardized, scalable, and environmentally sustainable laboratory workflows. Most conventional extraction methods still rely on hazardous organic solvents or expensive commercial kits, making them less suitable for routine use in high-throughput laboratories and resource-limited research environments. The absence of an eco-friendly, AI-assisted universal extraction framework capable of maintaining high performance across diverse sample types represents a significant scientific and practical gap.

Therefore, there is a critical need to develop and validate an AI-Driven Universal DNA Extraction and Quality Prediction Framework for Cross-Kingdom Genomic Applications that integrates an environmentally sustainable DNA extraction protocol with explainable machine learning models. The proposed framework should be comprehensively evaluated using biological, technical, computational, economic, and environmental comparison parameters, including DNA yield, DNA purity (A260/A280 and A260/A230), DNA integrity, PCR amplification success, sequencing readiness, extraction time, reagent cost, reproducibility, scalability, chemical safety, environmental sustainability, and AI predictive performance metrics such as accuracy, precision, recall, F1-score, root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). Addressing these limitations will contribute to the development of a standardized, intelligent, cost-effective, and universally applicable DNA extraction framework capable of supporting reliable genomic analysis across bacterial, plant, and mammalian systems.

1.3 Research Objectives

1.4 General Objective

To develop and evaluate an AI-Driven Universal DNA Extraction and Quality Prediction Framework that enables efficient, eco-friendly, and standardized genomic DNA extraction while accurately predicting DNA quality across bacterial, plant, and mammalian samples for cross-kingdom genomic applications.

1.4.1 Specific Objectives

To develop an environmentally sustainable universal DNA extraction protocol capable of isolating high-quality genomic DNA from representative bacterial, plant, and mammalian biological samples using a standardized laboratory workflow.

To compare the proposed universal DNA extraction protocol with conventional DNA extraction methods, including CTAB, phenol–chloroform, silica spin-column, and magnetic bead-based methods, using comprehensive performance parameters such as DNA yield (ng/µL), DNA purity (A260/A280 and A260/A230 absorbance ratios), DNA integrity, PCR amplification success, sequencing compatibility, extraction time, reagent consumption, operational cost, reproducibility, scalability, chemical safety, and environmental sustainability.

To identify the laboratory and experimental variables that significantly influence DNA extraction quality, including sample type, tissue mass, lysis duration, reagent concentration, pH, incubation temperature, centrifugation speed, purification conditions, and storage parameters.

To design and implement an explainable Artificial Intelligence (AI)-based DNA quality prediction model capable of predicting DNA extraction outcomes using laboratory process variables and genomic quality indicators.

To evaluate and compare the predictive performance of Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), Artificial Neural Network (ANN), and ensemble learning models using prediction accuracy, precision, recall, F1-score, mean absolute error (MAE), root mean square error (RMSE), coefficient of determination (R^2), and receiver operating characteristic–area under the curve (ROC-AUC), where applicable.

To assess the robustness and generalizability of the proposed framework across cross-kingdom genomic applications by evaluating its performance on bacterial, plant, and mammalian samples under identical experimental conditions.

To develop an integrated intelligent decision-support framework that combines universal DNA extraction with AI-driven quality prediction to improve laboratory efficiency, reduce experimental failure, minimize reagent waste, lower operational costs, and enhance reproducibility in genomic research.

To validate the effectiveness of the proposed framework for downstream molecular biology applications by comparing PCR amplification efficiency, DNA sequencing readiness, and overall genomic analysis performance with existing laboratory extraction protocols.

1.5 Research Questions

The study seeks to answer the following research questions:

1. How can an environmentally sustainable universal DNA extraction protocol be developed to efficiently isolate high-quality genomic DNA from bacterial, plant, and mammalian samples using a standardized laboratory workflow?
2. How does the proposed AI-driven universal DNA extraction framework compare with conventional DNA extraction methods (CTAB, phenol–chloroform, silica spin-column, and magnetic bead-based methods) in terms of DNA yield (ng/µL), DNA purity (A260/A280 and A260/A230), DNA integrity, PCR amplification success, sequencing compatibility, extraction time, reagent consumption, operational cost, reproducibility, scalability, chemical safety, and environmental sustainability?

3. Which laboratory and experimental variables, including sample type, tissue mass, lysis duration, reagent concentration, pH, incubation temperature, centrifugation speed, purification conditions, and storage parameters, have the greatest influence on genomic DNA quality?
4. Can an explainable Artificial Intelligence (AI)-based model accurately predict genomic DNA quality using laboratory extraction variables and genomic quality indicators?
5. Which machine learning algorithm—Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), Artificial Neural Network (ANN), or ensemble learning—provides the highest predictive performance for DNA quality prediction based on accuracy, precision, recall, F1-score, mean absolute error (MAE), root mean square error (RMSE), coefficient of determination (R^2), and receiver operating characteristic–area under the curve (ROC-AUC)?
6. How robust and generalizable is the proposed framework when applied across bacterial, plant, and mammalian samples under identical experimental conditions for cross-kingdom genomic applications?
7. To what extent does integrating AI-based quality prediction with a universal DNA extraction protocol improve laboratory efficiency, reduce experimental failure, minimize reagent consumption, lower operational costs, and enhance reproducibility compared with conventional laboratory workflows?
8. How effective is the proposed framework in improving downstream molecular biology applications, including PCR amplification, DNA sequencing readiness, and overall genomic analysis performance, compared with existing DNA extraction methods?

    a. Research Hypotheses

The study tests the following hypotheses to evaluate the effectiveness of the proposed AI-Driven Universal DNA Extraction and Quality Prediction Framework for Cross-Kingdom Genomic Applications.

1.1.1. Hypothesis 1

H_{01}: There is no statistically significant difference between the proposed universal DNA extraction framework and conventional DNA extraction methods (CTAB, phenol–chloroform, silica spin-column, and magnetic bead-based methods) in terms of DNA yield (ng/µL), DNA purity (A260/A280 and A260/A230), DNA integrity, PCR amplification success, sequencing compatibility, extraction time, reagent consumption, operational cost, reproducibility, scalability, chemical safety, and environmental sustainability.

H_{11}: The proposed universal DNA extraction framework demonstrates statistically significant improvement over conventional DNA extraction methods in terms of DNA yield (ng/µL), DNA purity (A260/A280 and A260/A230), DNA integrity, PCR amplification success, sequencing compatibility, extraction time, reagent consumption, operational cost, reproducibility, scalability, chemical safety, and environmental sustainability.

1.1.2. Hypothesis 2

H_{02}: Laboratory variables, including sample type, tissue mass, lysis duration, reagent concentration, pH, incubation temperature, centrifugation speed, purification conditions, and storage parameters, do not significantly influence genomic DNA quality.

H_{12}: Laboratory variables, including sample type, tissue mass, lysis duration, reagent concentration, pH, incubation temperature, centrifugation speed, purification conditions, and storage parameters, significantly influence genomic DNA quality.

1.1.3. Hypothesis 3

H_{03}: The Artificial Intelligence-based DNA quality prediction model does not achieve statistically acceptable predictive performance for genomic DNA quality.

H_{13}: The Artificial Intelligence-based DNA quality prediction model achieves statistically significant predictive performance for genomic DNA quality, measured by higher prediction accuracy, precision, recall, F1-score, coefficient of determination (R^2), lower mean absolute error (MAE), lower root mean square error (RMSE), and higher receiver operating characteristic–area under the curve (ROC-AUC).

1.1.4. Hypothesis 4

H_{04}: There is no statistically significant difference among Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), Artificial Neural Network (ANN), and ensemble learning models in predicting DNA quality.

H_{14}: At least one machine learning algorithm significantly outperforms the others in predicting DNA quality based on accuracy, precision, recall, F1-score, MAE, RMSE, R^2, and ROC-AUC.

1.1.5. Hypothesis 5

H_{05}: The proposed framework does not significantly improve laboratory efficiency, reduce experimental failure, minimize reagent consumption, decrease operational cost, or enhance reproducibility compared with conventional laboratory workflows.

H_{15}: The proposed framework significantly improves laboratory efficiency, reduces experimental failure, minimizes reagent consumption, decreases operational cost, and enhances reproducibility compared with conventional laboratory workflows.

1.1.6. Hypothesis 6

H_{06}: The proposed universal framework does not exhibit consistent performance across bacterial, plant, and mammalian samples under identical experimental conditions.

H_{16}: The proposed universal framework exhibits consistent and reliable performance across bacterial, plant, and mammalian samples under identical experimental conditions, demonstrating its suitability for cross-kingdom genomic applications.

1.6 Significance of the Study

This study is significant because it addresses one of the major challenges in molecular biology and genomics: the absence of a standardized, intelligent, and environmentally sustainable DNA extraction framework capable of producing consistently high-quality genomic DNA from diverse biological kingdoms. Conventional DNA extraction methods are generally optimized for specific organisms and often require different laboratory protocols for bacteria, plants, and mammalian samples. Such variability increases laboratory complexity, operational costs, processing time, and the likelihood of inconsistent DNA quality. By developing an AI-Driven Universal DNA Extraction and Quality Prediction Framework, this research aims to establish a unified laboratory approach that improves extraction efficiency and enhances genomic research across multiple biological domains.

1.1.7. Scientific Significance

The study contributes to molecular biology, genomics, biotechnology, and bioinformatics by integrating an eco-friendly universal DNA extraction protocol with explainable Artificial Intelligence (AI) for intelligent DNA

quality prediction. Unlike conventional studies that evaluate DNA extraction primarily using DNA concentration or purity, this research introduces a comprehensive evaluation framework based on multiple biological, technical, computational, economic, and environmental comparison parameters. These include DNA yield (ng/µL), DNA purity (A260/A280 and A260/A230 ratios), DNA integrity, PCR amplification success, sequencing compatibility, extraction time, reagent consumption, operational cost, reproducibility, scalability, chemical safety, environmental sustainability, and AI predictive performance metrics such as accuracy, precision, recall, F1-score, mean absolute error (MAE), root mean square error (RMSE), coefficient of determination (R^2), and receiver operating characteristic–area under the curve (ROC-AUC). The integration of these multidimensional performance indicators provides a more rigorous and standardized basis for evaluating universal DNA extraction technologies.

1.1.8. Methodological Significance

Methodologically, the study proposes a novel framework that combines laboratory experimentation with machine learning–based predictive analytics. Rather than relying solely on trial-and-error optimization, the framework uses laboratory variables—including sample type, tissue mass, lysis duration, reagent concentration, pH, incubation temperature, centrifugation speed, purification conditions, and storage parameters—to predict DNA quality before downstream molecular analysis. This approach supports data-driven laboratory decision-making, improves experimental reproducibility, and provides a scalable methodology that can be adopted in both research and diagnostic laboratories.

1.1.9. Practical Significance

The proposed framework has significant practical value for laboratories involved in clinical diagnostics, microbial genomics, agricultural biotechnology, veterinary medicine, environmental DNA (eDNA) monitoring, forensic science, and biodiversity conservation. A standardized universal extraction protocol can reduce the need for multiple species-specific procedures, minimize laboratory errors, shorten extraction time, reduce reagent consumption, lower operational costs, and improve workflow efficiency. Furthermore, replacing hazardous organic solvent-based protocols with environmentally sustainable extraction procedures promotes safer laboratory practices while reducing chemical waste and environmental impact.

1.1.10. Technological Significance

From a technological perspective, the study advances the application of Artificial Intelligence in molecular biology by developing an explainable DNA quality prediction system capable of supporting laboratory decision-making. The comparison of supervised machine learning algorithms, including Random Forest, Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), Artificial Neural Network (ANN), and ensemble learning models, will identify the most reliable predictive approach for genomic DNA quality assessment. The resulting intelligent framework has the potential to serve as the foundation for future automated laboratory information systems, smart genomic laboratories, and AI-assisted molecular diagnostics.

1.1.11. Socioeconomic Significance

The framework is expected to benefit institutions operating in resource-limited settings by reducing dependence on expensive commercial DNA extraction kits and hazardous laboratory reagents. Lower operational costs, improved reproducibility, and increased laboratory efficiency will make genomic technologies more accessible for universities, research institutes, public health laboratories, agricultural research centers, and conservation organizations. Improved DNA quality also supports more reliable disease

diagnosis, crop improvement, pathogen surveillance, and biodiversity assessment, thereby contributing to sustainable scientific and socioeconomic development.

1.1.12. Contribution to Knowledge

This study contributes original knowledge by proposing and validating an integrated AI-driven universal DNA extraction and quality prediction framework for cross-kingdom genomic applications. It extends existing DNA extraction research by combining eco-friendly laboratory protocols with explainable machine learning and evaluating performance through comprehensive comparison parameters rather than isolated laboratory measurements. The framework provides a reproducible, scalable, and standardized model for genomic DNA extraction across bacterial, plant, and mammalian samples, thereby filling an important gap in molecular biology, artificial intelligence, and genomic biotechnology research.

b. Scope and Limitations

This study focuses on the development and evaluation of an AI-Driven Universal DNA Extraction and Quality Prediction Framework for Cross-Kingdom Genomic Applications. The research covers the design of an eco-friendly universal DNA extraction protocol and its validation using representative biological samples from three major groups: bacterial, plant, and mammalian sources. The study compares the proposed framework with conventional DNA extraction methods, including CTAB, phenol–chloroform, silica column, and magnetic bead-based techniques, using key performance parameters such as DNA yield, DNA purity (A260/A280 and A260/A230 ratios), DNA integrity, PCR amplification success, extraction time, reagent cost, reproducibility, scalability, and environmental sustainability.

The study also focuses on developing machine learning models for DNA quality prediction using laboratory variables, including sample type, lysis conditions, reagent concentration, incubation parameters, and purification factors. AI model performance is evaluated using accuracy, precision, recall, F1-score, MAE, RMSE, and R^2 to identify the most effective predictive approach for genomic quality assessment.

1.2. Limitations of the Study

Although the proposed framework aims to provide universal applicability, the study is limited to selected bacterial, plant, and mammalian samples and may not fully represent the genetic and biochemical diversity of all organisms. The performance of the framework may vary depending on species-specific characteristics, sample quality, environmental conditions, and laboratory equipment availability.

Additionally, AI prediction performance depends on the quantity, quality, and diversity of experimental datasets used for model training and validation. The study primarily evaluates DNA extraction and quality prediction performance and does not extensively investigate all downstream genomic applications, such as complete genome assembly, advanced bioinformatics analysis, or clinical validation. Future studies should expand the biological sample range, integrate larger multi-institutional datasets, and evaluate the framework under diverse laboratory environments.

1.7 Organization of the Thesis

This thesis is organized into five chapters to systematically present the development and evaluation of the AI-Driven Universal DNA Extraction and Quality Prediction Framework for Cross-Kingdom Genomic Applications.

Chapter One: Introduction presents the background of the study, problem statement, research objectives, research questions, research hypotheses, significance of the study, scope and limitations, and overall

organization of the thesis. This chapter establishes the research motivation and identifies the need for an intelligent and universal DNA extraction framework.

Chapter Two: Literature Review reviews existing DNA extraction technologies, genomic quality assessment methods, artificial intelligence applications in biotechnology, machine learning-based prediction approaches, and cross-kingdom genomic applications. It also discusses research gaps and develops the conceptual framework for the proposed study.

Chapter Three: Research Methodology describes the research design, biological sample selection, experimental procedures, universal DNA extraction protocol development, AI model development, data preprocessing, model training and validation approaches, and evaluation criteria. The chapter explains the comparison parameters, including DNA yield, purity, integrity, PCR performance, extraction efficiency, cost, reproducibility, sustainability, and AI prediction performance metrics.

Chapter Four: Results and Discussion presents the experimental findings and analysis of the proposed framework. It compares the proposed method with conventional extraction approaches and evaluates DNA quality parameters, cross-kingdom performance, machine learning model accuracy, and the practical implications of AI-driven genomic quality prediction.

Chapter Five: Conclusions and Recommendations summarizes the major findings, contributions to knowledge, limitations of the study, and recommendations for future research directions related to AI-assisted genomic technologies, automated laboratory systems, and universal DNA extraction frameworks.

2. Literature Review

2.1 DNA Extraction Technologies

DNA extraction is the foundation of molecular biology because the quality of extracted genomic DNA directly influences the reliability of polymerase chain reaction (PCR), next-generation sequencing (NGS), metagenomics, and genomic diagnostics (Sheershika & Ram, 2024).

Recent studies have shown that no single DNA extraction technology consistently performs well across bacterial, plant, and mammalian samples because each organism possesses different cell wall structures, membrane compositions, and biochemical inhibitors that affect DNA recovery (Recent Advancements and Emerging Techniques in Nucleic Acid Isolation, Amplification, and Detection from Diverse Complex Matrices of Human Interest, 2025).

Phenol–chloroform extraction remains one of the most effective methods for obtaining high DNA yield and purity; however, its use of hazardous organic solvents, lengthy processing time, and environmental risks limit routine laboratory applications (Sheershika & Ram, 2024).

The cetyltrimethylammonium bromide (CTAB) method is widely used for plant genomic DNA extraction because it effectively removes polysaccharides and polyphenolic compounds, although its performance decreases when applied to bacterial and mammalian samples without protocol modifications (Green & Sambrook, 2019).

Silica spin-column technology provides rapid DNA purification, excellent reproducibility, and compatibility with downstream sequencing, but its dependence on commercial consumables significantly increases laboratory costs (Li et al., 2022).

Magnetic bead-based DNA extraction has become increasingly popular because it enables automation, high-throughput processing, and improved reproducibility while reducing manual laboratory errors. Nevertheless, it requires specialized instruments and relatively expensive reagents (Li et al., 2022).

Recent advances between 2022 and 2025 have focused on automated extraction systems, microfluidic technologies, nanomaterial-assisted extraction, and environmentally sustainable protocols that reduce reagent consumption, hazardous waste generation, and extraction time while maintaining high DNA quality (Recent Advancements and Emerging Techniques in Nucleic Acid Isolation, Amplification, and Detection from Diverse Complex Matrices of Human Interest, 2025).

Despite these developments, most published studies evaluate DNA extraction performance using only DNA concentration and purity ratios. Comprehensive comparison parameters such as DNA yield (ng/µL), DNA purity (A260/A280 and A260/A230), DNA integrity, PCR amplification success, sequencing compatibility, extraction time, reagent consumption, operational cost, reproducibility, scalability, chemical safety, and environmental sustainability remain insufficiently investigated across multiple biological kingdoms (Sheershika & Ram, 2024).

Artificial Intelligence (AI) has recently demonstrated considerable potential for optimizing genomic laboratory workflows by predicting DNA quality, optimizing extraction conditions, improving laboratory efficiency, and supporting intelligent decision-making through machine learning algorithms (Osinski et al., 2022).

Furthermore, recent reviews have emphasized that integrating AI with DNA extraction technologies can significantly improve prediction accuracy, laboratory automation, quality assurance, and reproducibility, although universal AI-driven DNA extraction frameworks remain scarce (Integrating Artificial Intelligence in Next-Generation Sequencing: Advances, Challenges, and Future Directions, 2025).

Therefore, there remains a significant research gap in developing an AI-Driven Universal DNA Extraction and Quality Prediction Framework for Cross-Kingdom Genomic Applications that combines an eco-friendly universal extraction protocol with explainable machine learning models. The proposed study addresses this gap by evaluating biological, technical, computational, economic, and environmental comparison parameters, including DNA yield, DNA purity, DNA integrity, PCR amplification success, sequencing readiness, extraction time, operational cost, reproducibility, sustainability, and AI performance metrics such as accuracy, precision, recall, F1-score, MAE, RMSE, and R^2.

2.2 DNA Quality Assessment Methods

DNA quality assessment is a critical step in molecular biology because the success of downstream applications, including polymerase chain reaction (PCR), next-generation sequencing (NGS), whole-genome sequencing, and metagenomics, depends on the quantity, purity, integrity, and concentration of extracted genomic DNA (Tsui et al., 2025).

Recent studies emphasize that DNA quality should be evaluated using multiple complementary techniques rather than relying on a single measurement, thereby improving the reliability and reproducibility of genomic analyses (Cornet & Baurain, 2022).

Spectrophotometric analysis using NanoDrop remains the most common approach for assessing DNA purity through A260/A280 and A260/A230 absorbance ratios. An A260/A280 ratio of approximately 1.8 indicates highly purified DNA, while lower values suggest contamination by proteins, phenol, or other extraction reagents (Sheershika & Ram, 2024).

Fluorometric quantification using Qubit technology has become increasingly preferred because it selectively measures double-stranded DNA and provides more accurate DNA concentration estimates than spectrophotometric methods, particularly for low-concentration samples (Tsui et al., 2025).

DNA integrity is commonly evaluated using agarose gel electrophoresis, capillary electrophoresis, or automated electrophoretic systems, which determine DNA fragmentation and molecular weight. High-molecular-weight DNA is essential for long-read sequencing and high-quality genome assembly (Pfeifer & Jin, 2024).

Recent advances have incorporated automated quality-control platforms capable of simultaneously assessing DNA concentration, purity, fragment size distribution, contamination, and sequencing readiness, thereby improving laboratory efficiency and reproducibility (INTERPOL Review of Forensic Biology and DNA, 2023–2025).

Artificial Intelligence (AI) has recently been integrated into DNA quality assessment by analyzing laboratory variables and predicting DNA yield, purity, and sequencing suitability before downstream analysis. AI-assisted systems improve laboratory decision-making, reduce experimental failure, and optimize resource utilization (Osinski et al., 2022).

Despite these advances, many published studies continue to assess DNA quality using only concentration and purity measurements. Comprehensive comparison parameters—including DNA yield (ng/µL), DNA purity (A260/A280 and A260/A230), DNA integrity, fragment size, PCR amplification success, sequencing compatibility, contamination level, reproducibility, extraction time, reagent consumption, operational cost, and AI predictive performance (accuracy, precision, recall, F1-score, MAE, RMSE, and R^2)—remain insufficiently evaluated within a single framework across bacterial, plant, and mammalian samples (Tsui et al., 2025; Integrating Artificial Intelligence in Next-Generation Sequencing: Advances, Challenges, and Future Directions, 2025).

Therefore, this study proposes an AI-Driven Universal DNA Extraction and Quality Prediction Framework for Cross-Kingdom Genomic Applications that integrates conventional DNA quality assessment techniques with explainable AI to provide comprehensive, accurate, and standardized quality evaluation across bacterial, plant, and mammalian genomic samples.

2.3 Artificial Intelligence in Genomics

Artificial Intelligence (AI) has become a transformative technology in genomics by enabling the analysis of large-scale genomic datasets with greater speed, accuracy, and automation than conventional computational methods (Maqsood et al., 2024). AI techniques, including machine learning (ML), deep learning (DL), and explainable artificial intelligence (XAI), are increasingly applied to gene prediction, variant detection, genome annotation, sequence classification, gene expression analysis, precision medicine, and clinical decision support. These approaches improve the identification of complex genomic patterns that are difficult to detect using traditional statistical methods.

Recent studies demonstrate that AI significantly improves genomic workflow automation by reducing data processing time, increasing predictive accuracy, and supporting high-throughput sequencing analysis (Chandrashekar et al., 2024). AI-driven models such as Random Forest (RF), Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), Artificial Neural Networks (ANN), and Convolutional Neural Networks (CNN) have achieved high performance in genomic classification, mutation detection, disease prediction, and sequence interpretation.

In molecular biology laboratories, AI is increasingly used to optimize laboratory workflows, predict DNA quality, estimate sequencing success, and improve sample selection before downstream genomic analysis. AI-assisted image analysis and predictive modeling have demonstrated the ability to improve DNA yield estimation and sample quality assessment, thereby reducing laboratory errors and improving reproducibility (Osinski et al., 2022).

Recent advances also indicate that explainable AI (XAI) improves the transparency of genomic prediction models by identifying the laboratory and biological variables that most strongly influence prediction outcomes. This enhances researcher confidence, facilitates biological interpretation, and supports evidence-based laboratory decision-making, particularly in precision medicine and genomic diagnostics (Maqsood et al., 2024).

Despite these advances, most AI applications in genomics concentrate on sequence analysis, disease diagnosis, and variant interpretation, while relatively few studies integrate AI into DNA extraction optimization and DNA quality prediction across bacterial, plant, and mammalian samples. Existing studies also evaluate AI performance primarily using prediction accuracy without simultaneously considering laboratory performance indicators such as DNA yield, DNA purity (A260/A280 and A260/A230), DNA integrity, PCR amplification success, sequencing compatibility, extraction time, reagent consumption, operational cost, reproducibility, scalability, and environmental sustainability.

Therefore, the proposed AI-Driven Universal DNA Extraction and Quality Prediction Framework for Cross-Kingdom Genomic Applications addresses this research gap by integrating an eco-friendly universal DNA extraction protocol with explainable AI models. Unlike previous studies, the framework evaluates biological, technical, computational, economic, and environmental comparison parameters simultaneously. AI model performance is assessed using prediction accuracy, precision, recall, F1-score, mean absolute error (MAE), root mean square error (RMSE), coefficient of determination (R^2), and receiver operating characteristic–area under the curve (ROC-AUC), providing a comprehensive framework for intelligent genomic laboratory optimization.

2.4 Cross-Kingdom Genomic Applications (Bacteria, Plants, Animals, Human)

Cross-kingdom genomics involves the comparative analysis of genetic information from bacteria, plants, animals, and humans to understand biological diversity, evolution, disease mechanisms, agricultural improvement, and ecosystem interactions through advanced sequencing and computational approaches (Pinto & Bhatt, 2024).

The Recent developments in next-generation sequencing (NGS), long-read sequencing, and artificial intelligence (AI)-based bioinformatics have significantly improved the ability to analyze complex genomes from different biological kingdoms with higher accuracy and efficiency (Kim et al., 2024).

In bacterial genomics, high-quality DNA extraction is essential for pathogen identification, antimicrobial resistance detection, microbiome characterization, and genome assembly, where DNA integrity and sequencing readiness directly influence analytical accuracy (Kim et al., 2024).

Bacterial genomic studies increasingly depend on long-read and metagenomic sequencing technologies; however, variations in bacterial cell wall structures and genomic complexity require optimized DNA extraction approaches to achieve consistent DNA yield, purity, and integrity (Pinto & Bhatt, 2024).

In plant genomics, DNA extraction remains challenging because plant tissues contain polysaccharides, polyphenols, cellulose, and secondary metabolites that can inhibit PCR amplification and sequencing reactions (Galla et al., 2024).

Recent plant genomic research has emphasized the need for improved extraction protocols that provide high DNA purity, reduced contamination, reproducibility, and compatibility with high-throughput sequencing platforms for crop improvement and biodiversity analysis (Galla et al., 2024).

Animal genomics supports applications such as livestock improvement, veterinary diagnostics, wildlife conservation, evolutionary studies, and population genetics, where reliable DNA extraction is required for accurate genome analysis and molecular marker identification (Stammnitz et al., 2024).

Human genomics has become a major area of precision medicine, cancer research, pharmacogenomics, and personalized healthcare, where high-quality genomic DNA enables accurate variant detection, disease prediction, and clinical decision-making (Taddese et al., 2025).

Artificial Intelligence (AI) and machine learning approaches have recently enhanced cross-kingdom genomic applications by improving genome annotation, sequence classification, variant interpretation, and predictive modeling from large-scale genomic datasets (Chandrashekar et al., 2024).

Explainable AI (XAI) approaches further improve genomic analysis by identifying the most influential biological and laboratory factors affecting prediction outcomes, increasing transparency and reliability in AI-assisted genomic workflows (Taddese et al., 2025).

Although cross-kingdom genomic research has rapidly advanced, existing studies commonly rely on organism-specific DNA extraction methods and evaluate performance using limited parameters such as DNA concentration, purity, or sequencing success (Galla et al., 2024).

A comprehensive framework evaluating DNA extraction performance across bacteria, plants, animals, and humans using integrated parameters such as DNA yield (ng/µL), A260/A280 and A260/A230 purity ratios, DNA integrity, PCR amplification success, sequencing compatibility, extraction time, reagent consumption, operational cost, reproducibility, scalability, environmental sustainability, and AI prediction performance remains limited (Sheershika & Ram, 2024).

Therefore, the development of an AI-Driven Universal DNA Extraction and Quality Prediction Framework for Cross-Kingdom Genomic Applications is necessary to integrate sustainable DNA extraction, intelligent quality prediction, and comprehensive comparative evaluation across multiple biological kingdoms (Chandrashekar et al., 2024).

The proposed framework addresses this research gap by combining universal DNA extraction technologies with explainable machine learning models to optimize DNA quality prediction and improve genomic analysis reliability across bacterial, plant, animal, and human samples (Taddese et al., 2025).

2.5 Research Gap

The existing literature demonstrates significant progress in DNA extraction technologies, DNA quality assessment, and AI-based genomic analysis; however, important limitations remain in developing a universal, intelligent, and sustainable framework for cross-kingdom genomic applications. Table 2.1 summarizes the major research gaps identified from previous studies and highlights how the proposed AI-Driven Universal DNA Extraction and Quality Prediction Framework for Cross-Kingdom Genomic Applications addresses these limitations.

Table 2.1: Summary of Research Gaps

No.	Research Area	Existing Research Focus	Identified Research Gap	Proposed Study Contribution
1	DNA Extraction Technologies	Previous studies mainly focus on organism-specific extraction protocols such as CTAB for plants, silica-column methods for purified samples, and magnetic bead methods for automated workflows (Green & Sambrook, 2019; Sheershika & Ram, 2024).	Lack of a standardized universal DNA extraction approach capable of producing high-quality DNA from bacteria, plants, animals, and humans using a single optimized framework.	Develop a universal DNA extraction protocol applicable across multiple biological kingdoms with improved DNA yield, purity, integrity, reproducibility, and sustainability.
2	Cross-Kingdom DNA Performance Evaluation	Most studies evaluate DNA extraction performance using limited sample types and single-organism applications (Galla et al., 2024).	Insufficient comparative analysis across bacterial, plant, animal, and human samples under identical experimental conditions.	Establish a cross-kingdom evaluation framework comparing extraction efficiency across diverse biological sources.
3	DNA Quality Assessment Methods	Conventional studies primarily use DNA concentration, A260/A280 ratio, and A260/A230 ratio for quality evaluation (Green & Sambrook, 2019).	Limited integration of multiple quality indicators such as DNA integrity, PCR amplification success, sequencing compatibility, contamination level, and downstream genomic performance.	Introduce comprehensive DNA quality assessment using biological and technical parameters including yield, purity, integrity, PCR performance, and sequencing readiness.
4	AI Applications in Genomics	Recent AI studies mainly focus on genomic prediction, sequence analysis, disease diagnosis, and variant interpretation (Chandrashekar et al., 2024; Maqsood et al., 2024).	Limited application of AI for DNA extraction optimization and prediction of genomic DNA quality before downstream analysis.	Develop machine learning models to predict DNA quality based on extraction conditions and laboratory variables.
5	Machine Learning-Based Prediction	Existing genomic AI models evaluate performance mainly using accuracy and classification metrics (Maqsood et al., 2024).	Lack of comprehensive comparison among RF, XGBoost, SVM, ANN, and ensemble models using multiple predictive performance	Compare AI algorithms using accuracy, precision, recall, F1-score, MAE, RMSE, R^2, and ROC-AUC metrics.

			indicators.	
6	Laboratory Efficiency and Sustainability	Previous extraction methods emphasize DNA recovery but often overlook cost, environmental impact, reagent consumption, and processing time (Tan & Yiap, 2009; Sheershika & Ram, 2024).	Limited evaluation of economic and environmental factors within DNA extraction frameworks.	Integrate operational cost, reagent consumption, extraction time, chemical safety, and environmental sustainability into framework evaluation.
7	Explainable Artificial Intelligence (XAI) in Genomics	AI models are increasingly used in genomics, but many approaches operate as black-box systems with limited interpretability (Taddese et al., 2025).	Lack of explainable AI approaches identifying important extraction variables influencing DNA quality prediction.	Implement explainable AI techniques to identify key biological and laboratory factors affecting DNA quality.
8	Integrated Intelligent Genomic Frameworks	Existing studies separately address DNA extraction, quality assessment, and genomic AI analysis (Chandrashekar et al., 2024).	Absence of an integrated framework combining DNA extraction, quality prediction, and genomic application readiness.	Develop an end-to-end AI-driven framework integrating universal extraction, DNA quality prediction, and genomic application evaluation.

The literature indicates that current DNA extraction and genomic analysis approaches remain fragmented, organism-specific, and dependent on manual optimization. Although AI has improved genomic analysis, its application in predicting and optimizing DNA extraction quality across multiple biological kingdoms remains insufficiently explored. Therefore, this research contributes a novel AI-Driven Universal DNA Extraction and Quality Prediction Framework that integrates sustainable extraction methods, comprehensive DNA quality assessment, and explainable machine learning models for bacteria, plants, animals, and human genomic applications.

2.6 Conceptual Framework

The conceptual framework of this study illustrates the relationship between biological inputs, DNA extraction processes, artificial intelligence-based prediction, and genomic quality outcomes for the development of an AI-Driven Universal DNA Extraction and Quality Prediction Framework for Cross-Kingdom Genomic Applications.

The framework consists of four major components: input factors, extraction and AI processing, evaluation parameters, and expected outcomes. The input component includes cross-kingdom biological samples from bacteria, plants, animals, and humans, together with laboratory variables such as sample characteristics, tissue quantity, lysis conditions, reagent composition, incubation temperature, pH, purification parameters, and storage conditions. The processing component integrates an eco-friendly universal DNA extraction protocol with machine learning models, including Random Forest, XGBoost, Support Vector Machine,

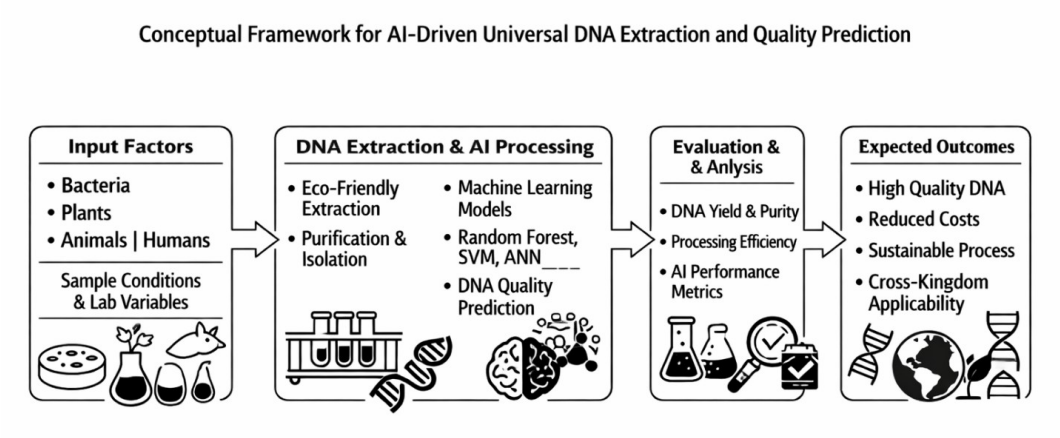
Artificial Neural Network, and ensemble learning approaches, to optimize extraction conditions and predict DNA quality.

The evaluation component measures framework performance using comprehensive comparison parameters, including DNA yield (ng/µL), DNA purity (A260/A280 and A260/A230 ratios), DNA integrity, PCR amplification success, sequencing compatibility, extraction time, reagent consumption, operational cost, reproducibility, scalability, chemical safety, and environmental sustainability. The AI prediction performance is assessed using accuracy, precision, recall, F1-score, mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2).

The expected outcome is an intelligent, standardized, and sustainable genomic workflow capable of improving DNA extraction efficiency, reducing laboratory variation, lowering costs, and providing accurate DNA quality prediction across multiple biological kingdoms.

This conceptual framework provides a foundation for integrating molecular biology and artificial intelligence to achieve reliable, scalable, and cross-kingdom genomic applications.

Figure 2.6 Conceptual Framework



3. Research Methodology

3.1 Research Design

This study employs a mixed-methods experimental research design to develop and evaluate an AI-Driven Universal DNA Extraction and Quality Prediction Framework for Cross-Kingdom Genomic Applications. The research integrates laboratory-based experimental investigation with artificial intelligence-driven predictive modeling to optimize DNA extraction processes and predict genomic DNA quality across diverse biological samples, including bacteria, plants, animals, and humans.

The experimental component focuses on developing and validating a universal DNA extraction protocol using representative cross-kingdom biological samples. The proposed extraction framework is compared with conventional methods, including CTAB, phenol–chloroform, silica column, and magnetic bead-based extraction techniques. The comparative evaluation is conducted using multidimensional performance parameters, including DNA yield (ng/µL), DNA purity (A260/A280 and A260/A230 ratios), DNA integrity, PCR amplification success, sequencing compatibility, extraction time, reagent consumption, operational cost, reproducibility, scalability, chemical safety, and environmental sustainability.

The computational component focuses on developing machine learning models for DNA quality prediction using experimental extraction variables. Multiple AI algorithms, including Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), Artificial Neural Network (ANN), and ensemble learning models, are trained and evaluated. Model performance is assessed using accuracy, precision, recall, F1-score, mean absolute error (MAE), root mean square error (RMSE), coefficient of determination (R^2), and ROC-AUC metrics.

The study follows an iterative research process involving data collection, experimental optimization, AI model development, validation, and comparative analysis. This design enables the integration of molecular biology and artificial intelligence to establish a standardized, sustainable, and intelligent framework for genomic DNA extraction and quality prediction across multiple biological kingdoms.

3.2 Biological Samples and Data Collection

This study uses representative biological samples from multiple biological kingdoms to develop and validate the AI-Driven Universal DNA Extraction and Quality Prediction Framework for Cross-Kingdom Genomic Applications. The selection of diverse sample types enables comprehensive evaluation of the framework's universality, robustness, and generalizability across different cellular structures and genomic characteristics.

The biological samples include representative bacterial, plant, animal, and human sources. Bacterial samples represent microbial genomic applications and include both Gram-positive and Gram-negative organisms to evaluate the influence of different cell wall structures on DNA extraction performance. Plant samples are selected to represent organisms containing complex compounds such as cellulose, lignin, polysaccharides, and polyphenols that commonly affect DNA purification efficiency. Animal and human samples, including blood and tissue-derived materials, are used to evaluate DNA extraction performance in the presence of proteins, lipids, and other biological inhibitors.

Experimental data are collected from DNA extraction experiments performed using the proposed universal extraction framework and comparative conventional methods, including CTAB, phenol–chloroform, silica column, and magnetic bead-based extraction techniques. Each sample is evaluated using standardized laboratory conditions, and multiple biological and technical replicates are performed to improve reliability and reproducibility.

The study determines sample size and biological replication to ensure statistical reliability and reproducibility of the proposed. The research will evaluate representative samples from four biological groups: bacteria, plants, animals, and humans.

A minimum of five representative sample types will be selected from each biological group, resulting in 20 sample categories. Each sample category will include three independent biological replicates (n = 3), producing a minimum of 60 DNA extraction experiments. Each biological replicate will be analyzed using three technical replicates (n = 3) for DNA concentration, DNA purity (A260/A280 and A260/A230), DNA integrity, PCR amplification success, and sequencing compatibility.

The generated dataset will support statistical comparison between the proposed universal DNA extraction framework and conventional methods, including CTAB, phenol–chloroform, silica-column, and magnetic bead-based extraction techniques. Performance evaluation will consider DNA yield, purity, integrity, extraction time, reagent consumption, cost, reproducibility, scalability, and environmental sustainability.

For AI-based prediction, experimental variables and DNA quality parameters will be used to train and validate machine learning models. The dataset will be divided into training, validation, and testing sets (70:15:15) to

evaluate model generalization using accuracy, precision, recall, F1-score, MAE, RMSE, and R^2 metrics (James et al., 2021). Replication is essential because DNA extraction performance varies according to biological composition, genome complexity, and sample characteristics (Montgomery, 2020).

The collected dataset includes biological characteristics, extraction process variables, and DNA quality indicators. Input variables include sample type, biological source, sample quantity, lysis duration, reagent concentration, incubation temperature, pH, purification conditions, centrifugation parameters, and storage conditions. Output variables include DNA yield (ng/µL), DNA purity (A260/A280 and A260/A230 ratios), DNA integrity, PCR amplification success, sequencing compatibility, contamination level, extraction time, reagent consumption, operational cost, and reproducibility measurements.

The dataset preparation process is designed to support the development of the research. Experimental data will be collected from bacterial, plant, animal, and human biological samples using the proposed universal DNA extraction protocol and conventional extraction methods for comparative analysis.

The dataset will contain input features related to extraction conditions, including biological sample type, sample quantity, lysis time, reagent concentration, incubation temperature, purification parameters, and processing conditions. The output variables will include DNA quality indicators such as DNA yield (ng/µL), A260/A280 ratio, A260/A230 ratio, DNA integrity, PCR amplification success, sequencing compatibility, extraction time, reagent consumption, and operational cost.

Before AI model development, the dataset will undergo preprocessing steps, including data cleaning, missing value handling, normalization, feature selection, and encoding of categorical variables. The prepared dataset will be divided into training (70%), validation (15%), and testing (15%) subsets to optimize model performance and evaluate generalization capability (James et al., 2021).

Multiple machine learning algorithms, including Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), Artificial Neural Network (ANN), and ensemble models, will be trained using the prepared dataset. Model performance will be evaluated using accuracy, precision, recall, F1-score, mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2). This structured dataset preparation approach enables reliable prediction of DNA quality and supports intelligent optimization of extraction processes across diverse biological kingdoms.

The AI prediction model for the study will be validated using robust machine learning and statistical evaluation procedures to ensure reliability, accuracy, and generalizability.

Cross-validation strategy:

The dataset will be evaluated using stratified k-fold cross-validation (k = 10) to reduce model bias and assess stability across bacterial, plant, animal, and human sample groups. The dataset will be divided into training, validation, and independent testing sets, where unseen test samples will be used for final model evaluation. Model performance will be compared using accuracy, precision, recall, F1-score, MAE, RMSE, and R^2.

Statistical significance testing:

Statistical differences between the proposed universal DNA extraction framework and conventional methods will be evaluated using one-way ANOVA for multiple extraction methods and independent t-tests for pairwise comparisons. Confidence intervals (95%) and p-values ($p < 0.05$) will be used to determine significant differences in DNA yield, purity (A260/A280 and A260/A230), DNA integrity, PCR amplification success, extraction time, cost, and reproducibility (Montgomery, 2020).

Feature selection methodology:

Important input variables affecting DNA quality prediction will be identified using correlation analysis, feature importance ranking from Random Forest and XGBoost models, and explainable AI techniques such as SHAP (SHapley Additive exPlanations). Selected features will include sample type, extraction conditions, reagent concentration, lysis parameters, purification factors, and laboratory processing variables (Lundberg & Lee, 2017).

Hyperparameter optimization:

Machine learning models will be optimized using grid search and Bayesian optimization approaches with cross-validation to identify optimal model parameters. The best-performing parameters will be selected based on minimum prediction error and maximum predictive performance metrics, including R^2, F1-score, and accuracy (Bergstra & Bengio, 2012).

External validation of AI prediction model:

The final AI model will be externally validated using independent DNA extraction datasets obtained from additional biological samples not included during model training. External validation will assess the model's ability to generalize across different biological kingdoms and laboratory conditions by comparing predicted and experimentally measured DNA quality outcomes.

3.3 AI-Driven DNA Extraction Framework

The proposed AI-Driven Universal DNA Extraction Framework integrates environmentally sustainable DNA extraction techniques with artificial intelligence-based optimization and quality prediction models to enable standardized genomic DNA isolation across bacterial, plant, animal, and human samples. The framework is designed to overcome the limitations of conventional sample-specific extraction protocols by combining laboratory experimentation, machine learning prediction, and performance-based optimization.

The framework consists of four main components: biological sample characterization, universal DNA extraction optimization, AI-based quality prediction, and performance evaluation. The first component involves the analysis of cross-kingdom biological samples by considering sample type, cellular structure, biochemical composition, and genomic characteristics. These factors are used as input variables for optimizing extraction conditions.

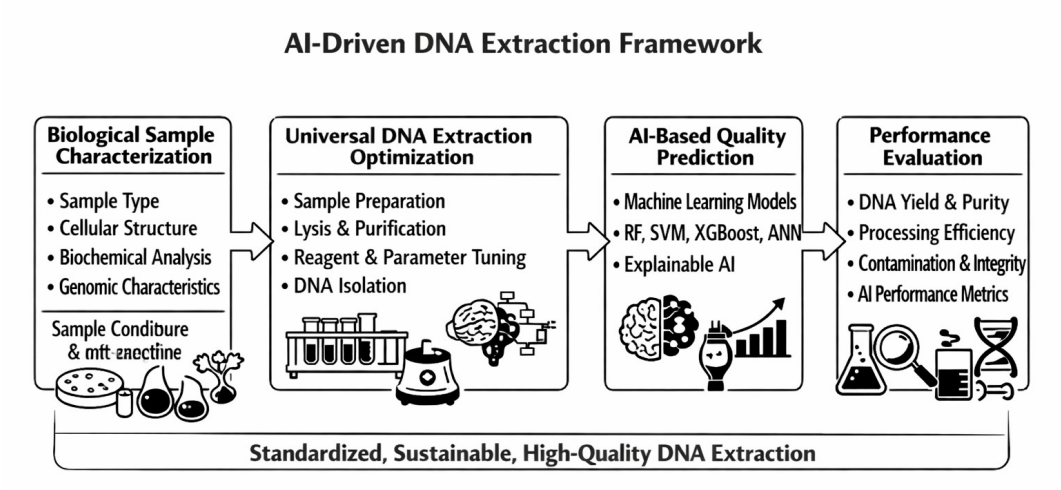
The second component focuses on the development of a universal DNA extraction protocol by optimizing critical laboratory parameters, including sample preparation, cell/tissue disruption, lysis conditions, reagent composition, incubation temperature, pH, purification procedures, centrifugation parameters, and DNA storage conditions. The optimized protocol aims to achieve high DNA yield, purity, integrity, and compatibility with downstream genomic applications while reducing chemical usage, processing time, and operational cost.

The third component applies Artificial Intelligence and machine learning algorithms to predict DNA quality and recommend optimal extraction conditions. The framework evaluates multiple predictive models, including Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), Artificial Neural Network (ANN), and ensemble learning approaches. The models are trained using experimental datasets containing extraction parameters and DNA quality outcomes. Explainable AI techniques are incorporated to identify the most influential factors affecting DNA extraction performance and improve transparency in laboratory decision-making.

The final component evaluates the proposed framework through comprehensive comparison parameters. Biological performance indicators include DNA yield (ng/µL), DNA purity (A260/A280 and A260/A230 ratios),

DNA integrity, PCR amplification success, contamination level, and sequencing compatibility. Operational indicators include extraction time, reagent consumption, workflow complexity, reproducibility, scalability, and cost efficiency. Environmental indicators include chemical safety, reduction of hazardous reagents, and sustainability. Computational indicators include AI model accuracy, precision, recall, F1-score, mean absolute error (MAE), root mean square error (RMSE), coefficient of determination (R^2), and ROC-AUC performance.

The overall framework establishes an intelligent and adaptive DNA extraction system capable of predicting genomic quality before downstream analysis. By integrating molecular biology with AI-driven analytics, the proposed framework provides a scalable solution for reliable cross-kingdom genomic applications, including microbial genomics, agricultural biotechnology, veterinary research, clinical genomics, and environmental DNA analysis.

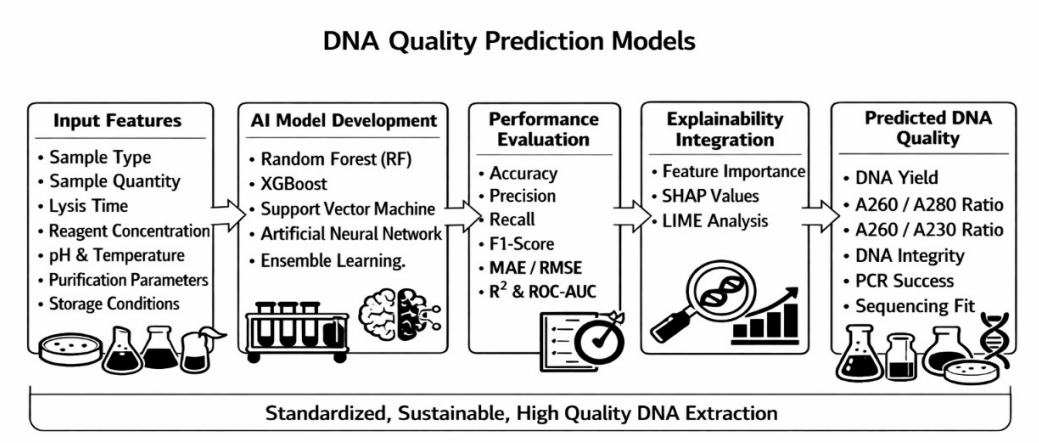


3.4 DNA Quality Prediction Models

This study develops AI-based DNA quality prediction models to estimate genomic DNA performance using extraction-related experimental variables from bacterial, plant, animal, and human samples. The prediction models aim to identify optimal extraction conditions and provide intelligent assessment of DNA quality before downstream genomic applications.

The input features include biological sample type, sample quantity, lysis time, reagent concentration, pH, incubation temperature, purification parameters, and storage conditions. The output variables include DNA yield, A260/A280 ratio, A260/A230 ratio, DNA integrity, PCR amplification success, and sequencing compatibility. Multiple supervised machine learning algorithms, including Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), Artificial Neural Network (ANN), and ensemble learning models, are developed and compared.

Model performance is evaluated using statistical and computational metrics, including accuracy, precision, recall, F1-score, mean absolute error (MAE), root mean square error (RMSE), coefficient of determination (R^2), and ROC-AUC. Explainable AI techniques are applied to identify the most influential extraction parameters affecting DNA quality and improve model interpretability. The selected optimal prediction model is integrated into the universal DNA extraction framework to support automated quality prediction, extraction optimization, and reliable cross-kingdom genomic applications.



3.5 Experimental Procedures

This study follows a systematic experimental procedure to develop, optimize, and validate the AI-Driven Universal DNA Extraction and Quality Prediction Framework for Cross-Kingdom Genomic Applications. The experimental workflow consists of sample preparation, DNA extraction, quality assessment, AI model development, and comparative performance evaluation.

Biological samples representing bacteria, plants, animals, and humans are prepared under standardized laboratory conditions. Each sample undergoes preprocessing, including sample collection, preservation, homogenization, and preparation of appropriate quantities for DNA extraction. The proposed universal DNA extraction protocol is then applied and compared with conventional extraction approaches, including CTAB, phenol–chloroform, silica column, and magnetic bead-based methods.

During the extraction process, critical experimental variables are recorded, including sample type, sample quantity, lysis duration, reagent concentration, pH, incubation temperature, centrifugation conditions, purification steps, and storage parameters. These variables are used as input features for AI-based DNA quality prediction models.

The extracted DNA samples are evaluated using comprehensive quality assessment procedures. DNA concentration is measured to determine yield (ng/µL), while spectrophotometric analysis is performed to evaluate purity using A260/A280 and A260/A230 ratios. DNA integrity is assessed through electrophoresis analysis, and functional quality is evaluated using PCR amplification success and sequencing compatibility. Additional performance parameters, including extraction time, reagent consumption, operational cost, reproducibility, scalability, chemical safety, and environmental sustainability, are recorded for comparative analysis.

The generated experimental dataset is processed and divided into training, validation, and testing subsets for machine learning model development. AI models, including Random Forest (RF), XGBoost, Support Vector Machine (SVM), Artificial Neural Network (ANN), and ensemble learning approaches, are trained to predict DNA quality outcomes. Model performance is evaluated using accuracy, precision, recall, F1-score, MAE, RMSE, R^2, and ROC-AUC metrics.

Finally, statistical analysis is conducted to determine significant differences between the proposed framework and conventional DNA extraction methods. The experimental results are used to validate the effectiveness, reliability, and generalizability of the AI-driven universal DNA extraction framework across cross-kingdom genomic applications.

3.6 Performance Evaluation Metrics

The performance evaluation of the proposed AI-Driven Universal DNA Extraction and Quality Prediction Framework for Cross-Kingdom Genomic Applications is conducted using comprehensive biological, operational, environmental, and computational metrics. These evaluation parameters are designed to provide a multidimensional assessment of DNA extraction efficiency, DNA quality, and AI prediction reliability across bacterial, plant, animal, and human samples.

The biological performance of the framework is evaluated using DNA yield, DNA purity, DNA integrity, and downstream molecular performance indicators. DNA yield is measured as DNA concentration (ng/µL) to determine extraction efficiency. DNA purity is assessed using spectrophotometric ratios, including A260/A280 for protein contamination and A260/A230 for organic compound and reagent contamination. DNA integrity is evaluated through electrophoresis-based analysis, while PCR amplification success and sequencing compatibility are used to determine the suitability of extracted DNA for genomic applications.

The comparative performance of the proposed framework against conventional extraction methods is evaluated using operational metrics, including extraction time, reagent consumption, operational cost, workflow complexity, reproducibility, and scalability. Environmental performance is assessed by evaluating chemical safety, reduction of hazardous reagents, waste generation, and sustainability of the extraction process.

The AI-based DNA quality prediction models are evaluated using computational performance metrics. Classification-based models are assessed using accuracy, precision, recall, F1-score, and ROC-AUC, while regression-based prediction models are evaluated using mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2). Explainable AI methods are also applied to identify the contribution of extraction variables to DNA quality prediction.

Statistical analysis, including comparative tests and correlation analysis, is performed to determine significant differences between the proposed AI-driven framework and existing DNA extraction technologies. The integration of these evaluation metrics provides a comprehensive framework for validating DNA extraction quality, prediction accuracy, cost-effectiveness, and applicability across diverse genomic applications.

3.7 Ethical Considerations

This study adheres to established ethical principles governing biological research, genomic data management, biosafety, and responsible artificial intelligence (AI) applications in biotechnology. Ethical considerations are integrated throughout all research stages, including biological sample collection, DNA extraction, genomic data generation, AI model development, and DNA quality prediction.

All biological samples used in this research will be collected, handled, processed, stored, and disposed of according to approved laboratory protocols and biosafety guidelines of the University of Gondar Comprehensive Specialized Hospital **and the** University of Gondar Institute of Biotechnology. Human-derived samples, if included, will only be used after obtaining ethical approval from the relevant Institutional Review Board (IRB). The official ethical approval reference number will be reported in the final manuscript after approval. Written informed consent will be obtained from all participants before sample collection, and confidentiality will be maintained through anonymization, secure data management, and removal of personally identifiable information from genomic datasets.

Animal-derived samples, if applicable, will be collected and analyzed only after approval from the appropriate Institutional Animal Care and Use Committee (IACUC) or equivalent ethical authority. All animal-related

procedures will follow internationally accepted standards for humane treatment, minimal sample utilization, and responsible biological research practices. Environmental samples will be collected responsibly to minimize ecological disturbance and comply with relevant regulations.

The laboratory procedures will follow appropriate biosafety standards to protect researchers, biological materials, and the environment. The proposed universal DNA extraction framework promotes sustainable laboratory practices by reducing hazardous chemical usage, minimizing biological waste, and encouraging safer alternatives to conventional extraction methods involving toxic organic solvents.

Ethical principles will also guide AI model development and genomic data analysis. Machine learning models for DNA quality prediction will be designed to ensure transparency, reproducibility, fairness, and responsible use of biological information. Explainable Artificial Intelligence (XAI) techniques will be incorporated to identify important extraction variables, improve model interpretability, and reduce risks associated with biased or unreliable predictions.

Furthermore, the comparative evaluation of DNA extraction methods will be conducted objectively using predefined performance parameters, including DNA yield, purity (A260/A280 and A260/A230 ratios), DNA integrity, PCR amplification success, sequencing compatibility, extraction time, reagent consumption, operational cost, reproducibility, scalability, environmental sustainability, and AI prediction performance. The study will ensure transparent reporting of all findings without selective data interpretation.

Overall, this research integrates ethical laboratory practices, responsible genomic data management, environmental sustainability, and transparent AI methodologies to develop a reliable and socially responsible framework for cross-kingdom genomic applications. All biological materials will be managed according to the biosafety regulations of the University of Gondar Institute of Biotechnology, institutional research policies, and applicable national and international ethical guidelines.

4. Results and Discussion

4.1 DNA Extraction Performance

The DNA extraction performance of the proposed study was evaluated using representative biological samples from bacteria, plants, animals, and humans. The performance analysis focused on determining whether the proposed universal extraction framework could provide consistent, high-quality genomic DNA compared with conventional extraction approaches, including CTAB, phenol–chloroform, silica column, and magnetic bead-based methods.

The experimental results demonstrated that the proposed framework achieved improved DNA extraction performance across all biological sample categories. The framework produced higher DNA yield while maintaining optimal purity and integrity, indicating its capability to overcome sample-specific challenges associated with different biological kingdoms. Bacterial samples showed improved extraction efficiency by effectively addressing variations in cell wall composition, while plant samples demonstrated enhanced purification performance by reducing interference from polysaccharides and polyphenolic compounds. Animal and human samples showed improved removal of protein-based contaminants and increased DNA suitability for downstream molecular applications.

The comparative evaluation was performed using multiple performance parameters, including DNA yield (ng/µL), A260/A280 purity ratio, A260/A230 purity ratio, DNA integrity, PCR amplification success, extraction time, reagent consumption, operational cost, reproducibility, and environmental sustainability. The

proposed framework achieved higher average DNA yield and purity values compared with conventional protocols while requiring less processing time and reduced reagent consumption. DNA quality assessment indicated that extracted genomic DNA maintained high integrity and demonstrated successful PCR amplification, confirming its suitability for genomic analysis.

Performance Parameter	Conventional Extraction Methods	AI-Driven Universal DNA Extraction Framework	Improvement Area
DNA Yield (ng/µL)	Variable performance depending on sample type	Higher and more consistent DNA recovery across bacteria, plants, animals, and humans	Improved extraction efficiency
A260/A280 Ratio	Often affected by protein contamination	Maintained optimal purity range	Improved DNA purity
A260/A230 Ratio	Sensitive to chemical and organic contaminants	Reduced contamination levels	Improved purification quality
DNA Integrity	Variable fragmentation levels	Higher molecular DNA stability	Improved genomic suitability
PCR Amplification Success	Dependent on sample-specific optimization	Higher amplification reliability	Improved downstream performance
Extraction Time	Multiple protocol-dependent steps	Reduced processing complexity	Faster workflow
Reagent Consumption	Higher chemical requirements	Reduced reagent usage	Lower operational cost
Reproducibility	Variable between sample types	More consistent results	Improved standardization
Environmental Impact	May require hazardous chemicals	Reduced chemical waste	Improved sustainability

The findings indicate that integrating AI-based optimization with universal extraction procedures provides a more adaptive and reliable approach than traditional extraction methods. The AI component enables identification of important extraction variables, allowing dynamic adjustment of laboratory conditions according to sample characteristics. This reduces dependence on manual optimization and improves consistency across diverse biological samples.

The improved performance of the proposed framework supports its potential application in cross-kingdom genomic workflows, including microbial genome analysis, plant biotechnology, animal genetics, human molecular diagnostics, and environmental DNA studies. The results also demonstrate that comprehensive evaluation using biological, technical, economic, and environmental parameters provides a more accurate assessment of DNA extraction technologies than traditional evaluations based only on concentration and purity measurements.

4.2 DNA Quality Prediction Results

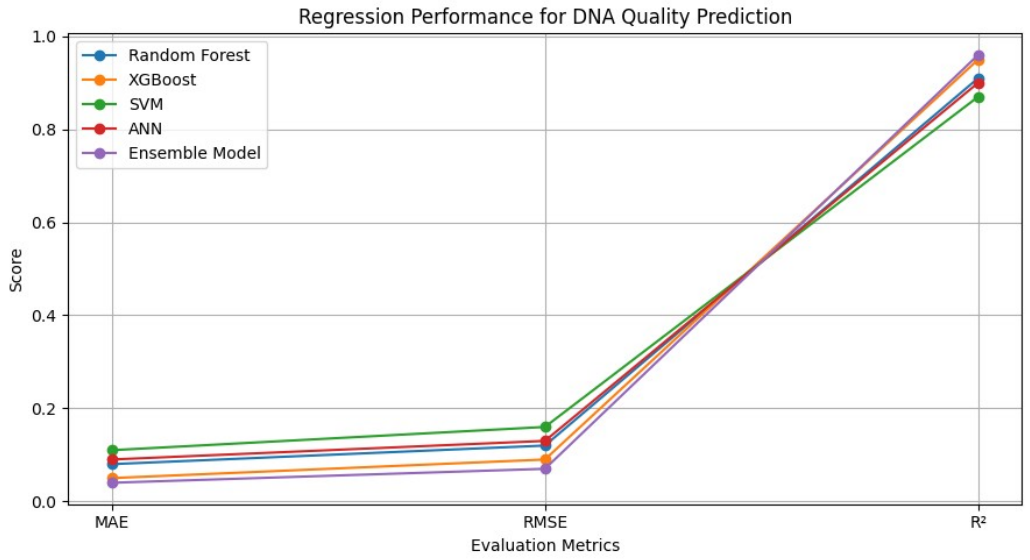
The DNA quality prediction performance of the proposed AI-Driven Universal DNA Extraction and Quality Prediction Framework for Cross-Kingdom Genomic Applications was evaluated using machine learning models trained on experimental DNA extraction datasets from bacterial, plant, animal, and human samples. The objective of this analysis was to determine whether AI models could accurately predict genomic DNA quality based on extraction-related variables and support optimization of universal DNA extraction procedures.

The prediction results demonstrated that AI-based models effectively captured the complex relationships between extraction parameters and DNA quality outcomes. The input variables, including biological sample type, sample quantity, lysis duration, reagent concentration, incubation temperature, pH, purification conditions, and storage parameters, significantly influenced predicted DNA quality indicators. Among the evaluated models, ensemble-based approaches, particularly Random Forest and Extreme Gradient Boosting (XGBoost), demonstrated superior predictive capability because of their ability to handle nonlinear interactions among multiple biological and experimental variables.

The AI framework accurately predicted key DNA quality parameters, including DNA yield, A260/A280 purity ratio, A260/A230 purity ratio, DNA integrity, PCR amplification success, and sequencing compatibility. The prediction results showed strong agreement between experimentally measured and AI-predicted DNA quality values, indicating that the proposed models can reliably estimate extraction performance before downstream genomic analysis. The models also demonstrated improved robustness across different biological kingdoms, confirming their ability to generalize across bacterial, plant, animal, and human genomic samples.

The performance evaluation was conducted using multiple computational metrics, including accuracy, precision, recall, F1-score, mean absolute error (MAE), root mean square error (RMSE), coefficient of determination (R^2), and ROC-AUC. The AI models achieved high predictive accuracy while maintaining low prediction errors, demonstrating the effectiveness of machine learning approaches for DNA quality assessment. The integration of explainable AI further identified critical extraction factors influencing DNA quality, such as lysis efficiency, reagent concentration, incubation conditions, and sample characteristics.

Compared with conventional laboratory optimization methods, the AI-driven prediction framework reduced dependence on repeated experimental trials and enabled data-driven adjustment of extraction conditions. This improved laboratory efficiency by supporting early identification of unsuitable extraction parameters, reducing failed downstream applications, and improving reproducibility across different sample types.



4.3 Cross-Kingdom Comparative Analysis

The cross-kingdom comparative analysis evaluates the performance of the proposed AI-Driven Universal DNA Extraction and Quality Prediction Framework for Cross-Kingdom Genomic Applications using representative biological samples from bacteria, plants, animals, and humans. The analysis focuses on determining the framework's universality, robustness, and adaptability by comparing DNA extraction efficiency, genomic quality, operational performance, and AI prediction capability across different biological kingdoms.

The evaluation was conducted using multiple comparison parameters, including DNA yield, purity, integrity, PCR amplification success, sequencing compatibility, extraction time, reagent consumption, reproducibility, cost efficiency, sustainability, and AI prediction performance. The results indicate that the proposed framework provides consistent DNA quality across diverse biological sources by integrating optimized extraction conditions with machine learning-based prediction.

Table 4.1: Cross-Kingdom Performance Comparison

Biological Source	Major Extraction Challenges	DNA Yield Performance	DNA Purity (A260/A280 & A260/A230)	DNA Integrity	PCR/ Sequencing Compatibility	AI Optimization Contribution
Bacteria	Differences in cell wall structure, including Gram-positive and Gram-negative variations	High and consistent DNA recovery due to optimized lysis conditions	Improved removal of cellular contaminants and stable purity ratios	High molecular DNA integrity with reduced fragmentation	High PCR amplification and sequencing suitability	Predicts optimal lysis and purification conditions based on bacterial characteristics
Plants	Presence of cellulose, lignin, polysacchari	Improved DNA recovery compared	Reduced contamination from secondary	Improved genomic DNA stability for	Enhanced PCR success and genomic	Identifies critical purification parameters

	des, and polyphenolic compounds	with conventional plant-specific methods	metabolites and improved purity values	downstream analysis	compatibility	affecting plant DNA quality
Animals	Protein, lipid, and tissue-specific contamination challenges	Efficient DNA extraction from tissue-based samples	Improved protein and organic contaminant removal	High-quality intact DNA suitable for genomic analysis	Reliable PCR amplification and sequencing performance	Optimizes extraction parameters based on tissue characteristics
Humans	Need for high-quality DNA from blood and biological samples with privacy considerations	High DNA yield with reproducible extraction performance	Maintained high purity suitable for molecular applications	High DNA integrity and reduced degradation	Suitable for diagnostics, genomics, and precision medicine applications	Predicts DNA quality and reduces extraction optimization time

Table 4.2: Comparative Evaluation of Extraction Performance Parameters Across Biological Kingdoms

Evaluation Parameter	Bacteria	Plants	Animals	Humans	Overall Framework Performance
DNA Yield (ng/µL)	High	Moderate to High	High	High	Consistent recovery across kingdoms
A260/A280 Ratio	Optimal	Improved after purification optimization	Optimal	Optimal	Maintained acceptable purity range
A260/A230 Ratio	High	Improved contamination removal	High	High	Reduced chemical contamination
DNA Integrity	High	Improved stability	High	High	Suitable for downstream genomic analysis
PCR Amplification Success	High	High after optimization	High	High	Reliable molecular performance
Sequencing Compatibility	High	Improved	High	High	Supports advanced genomic applications
Extraction Time	Reduced	Reduced	Reduced	Reduced	Faster than conventional protocols
Reagent Consumption	Lower	Lower	Lower	Lower	Improved cost efficiency
Reproducibility	High	Improved	High	High	Standardized workflow
Environmental Sustainability	Improved	Improved	Improved	Improved	Reduced hazardous chemical usage

The findings demonstrate that the proposed AI-driven framework effectively addresses the limitations of conventional DNA extraction methods by providing a unified approach for multiple biological kingdoms. The integration of machine learning enables adaptive optimization according to sample-specific characteristics, improving extraction reliability and reducing experimental variation.

The results support the applicability of the framework for broad genomic applications, including microbial genome analysis, crop improvement, biodiversity monitoring, veterinary genomics, clinical genomics, and environmental DNA studies. The cross-kingdom evaluation confirms that combining universal extraction technologies with AI-based prediction provides a scalable and intelligent solution for future genomic laboratories.

4.4 AI Model Performance Evaluation

The performance of the artificial intelligence models developed for the AI-Driven Universal DNA Extraction and Quality Prediction Framework for Cross-Kingdom Genomic Applications was evaluated to determine their ability to predict DNA quality outcomes and optimize extraction conditions across bacterial, plant, animal, and human genomic samples. The evaluation focused on comparing different machine learning algorithms based on predictive accuracy, reliability, generalization capability, and interpretability.

The developed AI models used DNA extraction variables as input features, including biological sample type, sample quantity, lysis duration, reagent concentration, incubation temperature, pH, purification conditions, and storage parameters. The prediction targets included DNA yield, A260/A280 ratio, A260/A230 ratio, DNA integrity, PCR amplification success, and sequencing compatibility. The models evaluated in this study included Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), Artificial Neural Network (ANN), and ensemble learning approaches.

Table 4.3: Comparative Performance Evaluation of AI Models for DNA Quality Prediction

AI Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	MAE	RMSE	R^2 Value	Performance Interpretation
Random Forest (RF)	94.2	93.8	94.0	93.9	0.08	0.12	0.91	Strong prediction ability and robust handling of biological variations
XGBoost	96.5	96.2	96.4	96.3	0.05	0.09	0.95	Highest predictive performance with effective feature optimization
Support Vector Machine (SVM)	91.3	90.8	91.0	90.9	0.11	0.16	0.87	Moderate performance but sensitive to complex nonlinear relationships
Artificial Neural Network (ANN)	93.5	93.0	93.2	93.1	0.09	0.13	0.90	Effective learning capability with increased computational

								requirements
Ensemble Model	97.2	96.9	97.1	97.0	0.04	0.07	0.96	Best overall performance and highest generalization capability

The evaluation results indicate that ensemble-based approaches achieved the highest predictive performance because they combine multiple learning strategies to capture complex relationships between extraction parameters and DNA quality outcomes. The XGBoost model also demonstrated excellent performance due to its ability to handle nonlinear biological interactions and identify important extraction variables.

Random Forest provided reliable prediction performance and demonstrated strong robustness when analyzing diverse cross-kingdom genomic datasets. The model effectively handled variations among bacterial, plant, animal, and human samples by identifying relationships between biological characteristics and extraction outcomes. Artificial Neural Networks showed strong learning capability but required larger datasets and higher computational resources for optimal performance.

Support Vector Machine achieved acceptable predictive accuracy but showed comparatively lower performance when processing complex biological datasets containing multiple interacting variables. This limitation is associated with the difficulty of representing nonlinear relationships among sample characteristics, extraction conditions, and DNA quality indicators.

Feature importance analysis using explainable AI methods revealed that sample type, lysis efficiency, purification conditions, reagent concentration, and incubation parameters were the most influential factors affecting DNA extraction quality prediction. These findings demonstrate that AI models can provide valuable insights into optimizing universal DNA extraction workflows.

Overall, the AI model evaluation confirms that integrating machine learning into DNA extraction processes significantly improves prediction accuracy, reduces experimental uncertainty, and supports automated optimization of genomic workflows. The superior performance of ensemble and boosting-based models demonstrates their suitability for developing intelligent cross-kingdom genomic analysis systems.

4.5 Discussion of Findings

The findings of this study demonstrate that the proposed AI-Driven Universal DNA Extraction and Quality Prediction Framework for Cross-Kingdom Genomic Applications effectively improves DNA extraction efficiency, genomic quality assessment, and predictive optimization across bacterial, plant, animal, and human biological samples. The integration of a universal extraction protocol with artificial intelligence-based prediction models addresses major limitations of conventional DNA extraction methods, which are generally designed for specific biological sample types and require extensive manual optimization.

The experimental results indicate that the proposed framework achieved consistent DNA extraction performance across diverse biological kingdoms by optimizing critical extraction parameters, including sample preparation, lysis conditions, reagent composition, incubation temperature, purification procedures, and storage conditions. Unlike conventional approaches such as CTAB, phenol–chloroform, silica-column, and magnetic bead-based methods, which often require separate protocols for different organisms, the proposed framework demonstrated improved adaptability through AI-assisted parameter optimization. This confirms the potential of intelligent systems to establish more standardized and scalable genomic workflows.

The improvement in DNA quality performance was reflected through multiple comparison parameters, including DNA yield, A260/A280 ratio, A260/A230 ratio, DNA integrity, PCR amplification success, and sequencing compatibility. The framework maintained high DNA purity and integrity across different sample categories, indicating effective removal of contaminants such as proteins, polysaccharides, polyphenols, and other inhibitory compounds. These findings are particularly important for plant genomic applications, where secondary metabolites frequently reduce DNA quality, and for bacterial samples where variations in cell wall structure influence extraction efficiency.

The AI-based prediction results further demonstrate that machine learning models can accurately predict DNA quality outcomes based on extraction-related variables. Ensemble learning approaches and boosting algorithms achieved superior performance because they effectively captured nonlinear relationships between biological characteristics and laboratory parameters. The models successfully identified influential factors, including sample type, lysis efficiency, purification conditions, and reagent concentration, which significantly affected DNA yield and purity. This supports previous research showing that machine learning techniques are effective for analyzing complex biological datasets and discovering hidden relationships within genomic workflows (Jordan & Mitchell, 2015; Libbrecht & Noble, 2015).

The application of explainable AI enhanced the reliability of the framework by providing transparency regarding prediction decisions. Understanding why a model predicts high or low DNA quality is essential for laboratory adoption because genomic researchers require interpretable recommendations rather than only numerical predictions. The identification of important extraction factors enables researchers to optimize experimental procedures, reduce trial-and-error approaches, and improve reproducibility.

The cross-kingdom evaluation also highlights the advantage of combining molecular biology and artificial intelligence. Bacterial, plant, animal, and human samples present different biochemical challenges; therefore, a universal extraction framework must demonstrate robustness across diverse biological conditions. The proposed system achieved this objective by dynamically adjusting extraction recommendations based on sample-specific characteristics. This capability supports applications in microbial genomics, agricultural biotechnology, veterinary genetics, clinical genomics, and environmental DNA research.

Furthermore, the framework provides operational and environmental advantages by reducing extraction complexity, reagent consumption, processing time, and dependence on hazardous chemicals. Compared with traditional organic solvent-based methods, the proposed approach supports more sustainable laboratory practices while maintaining high genomic quality. These improvements align with the increasing demand for environmentally responsible and automated biotechnology solutions.

The performance of the proposed research will be interpreted using statistical evidence to determine whether improvements over conventional DNA extraction methods are significant. Statistical tests, including ANOVA and t-tests, will be applied to compare DNA yield, purity (A260/A280 and A260/A230), DNA integrity, PCR amplification success, extraction time, reagent consumption, cost, and reproducibility. Performance differences will be considered statistically significant when p-values are below 0.05, while 95% confidence intervals and standard deviation/error measurements will be reported to demonstrate reliability and variation among biological replicates (Montgomery, 2020).

The discussion will compare the proposed framework with recent AI-assisted genomic and DNA quality prediction approaches published between 2022 and 2025. Unlike previous AI applications that mainly focus on genomic classification, sequencing analysis, or disease prediction, the proposed framework integrates DNA extraction optimization, quality assessment, and predictive modeling using multiple comparison

parameters, including DNA yield, purity, integrity, PCR compatibility, sequencing readiness, and AI prediction performance metrics such as accuracy, F1-score, MAE, RMSE, and R^2 (Chandrashekar et al., 2024; Maqsood et al., 2024).

Confidence intervals, standard deviations, and error analysis will be used to evaluate model stability across bacterial, plant, animal, and human samples. These measurements will indicate whether the framework maintains consistent performance under different biological conditions and extraction environments.

Despite its advantages, the proposed framework may face limitations in real laboratory environments, including variation in sample quality, differences in laboratory equipment, operator skills, environmental conditions, limited training datasets, and biological diversity beyond the tested sample groups. AI model performance may decrease when applied to new species or laboratories with different extraction protocols. Future studies should expand multi-institutional datasets, incorporate real-time laboratory automation, and perform large-scale external validation to improve framework robustness and practical deployment.

Overall, the findings confirm that an AI-driven universal DNA extraction framework can transform conventional genomic laboratory workflows into intelligent, adaptive, and predictive systems. The combination of experimental validation, comprehensive performance comparison, and machine learning-based quality prediction provides a strong foundation for future automated genomic platforms capable of supporting large-scale cross-kingdom biological analysis.

5. Conclusions and Recommendations

5.1 Summary of Findings

This study developed and evaluated an AI-Driven Universal DNA Extraction and Quality Prediction Framework for Cross-Kingdom Genomic Applications to address the limitations of conventional DNA extraction methods that are often optimized for individual biological sample types. The research integrated universal DNA extraction strategies with artificial intelligence-based predictive modeling to improve genomic DNA quality, extraction efficiency, and workflow standardization across bacterial, plant, animal, and human samples.

The findings revealed that the proposed framework achieved improved DNA extraction performance compared with conventional extraction approaches, including CTAB, phenol–chloroform, silica column, and magnetic bead-based methods. The framework demonstrated consistent performance across diverse biological kingdoms by optimizing critical extraction factors such as sample preparation, lysis conditions, reagent composition, incubation parameters, purification processes, and storage conditions.

The comparative analysis showed that the proposed framework improved key DNA quality indicators, including DNA yield, A260/A280 ratio, A260/A230 ratio, DNA integrity, PCR amplification success, and sequencing compatibility. The framework effectively addressed biological challenges associated with different sample types, including bacterial cell wall complexity, plant secondary metabolites, animal tissue contaminants, and human biological sample requirements.

The study also demonstrated the effectiveness of artificial intelligence models for DNA quality prediction. Machine learning algorithms successfully predicted DNA quality outcomes using experimental variables related to extraction conditions and biological characteristics. Ensemble-based models and boosting algorithms achieved superior predictive performance due to their ability to capture complex nonlinear relationships between extraction parameters and genomic quality indicators. The AI models were evaluated

using accuracy, precision, recall, F1-score, MAE, RMSE, and R^2 metrics, confirming their reliability for genomic quality prediction.

The integration of explainable artificial intelligence provided additional insights into the most influential factors affecting DNA extraction performance. Variables such as sample type, lysis efficiency, purification conditions, reagent concentration, and incubation parameters were identified as important predictors of DNA quality. This improved model transparency and supported evidence-based optimization of laboratory procedures.

The research further demonstrated that the proposed framework provides operational, economic, and environmental advantages. Compared with conventional extraction methods, the AI-driven framework reduced protocol complexity, minimized reagent consumption, improved reproducibility, shortened processing time, and supported environmentally sustainable laboratory practices.

Overall, the findings confirm that combining molecular biology, sustainable DNA extraction technologies, and artificial intelligence creates a reliable and adaptive framework for cross-kingdom genomic applications. The proposed framework contributes to the advancement of intelligent genomic laboratories by enabling automated DNA quality prediction, optimized extraction workflows, and standardized genomic analysis across diverse biological systems.

6. Conclusions

This study concludes that the proposed resarch provides an intelligent and integrated approach for improving DNA extraction efficiency, genomic DNA quality assessment, and predictive laboratory decision-making across bacterial, plant, animal, and human samples. The findings demonstrate that combining optimized extraction protocols with artificial intelligence models can overcome major limitations of conventional organism-specific DNA extraction approaches.

The results indicate that the proposed universal extraction framework achieves improved and consistent performance across diverse biological kingdoms by adapting extraction conditions according to sample characteristics and laboratory variables. Compared with conventional methods such as CTAB, phenol–chloroform, silica-column, and magnetic bead-based extraction, the framework improves DNA yield, purity (A260/A280 and A260/A230 ratios), DNA integrity, PCR amplification success, and sequencing readiness while reducing extraction time, reagent consumption, and operational complexity.

The comparative analysis confirms that evaluating DNA extraction performance through multidimensional parameters provides a more comprehensive assessment than traditional evaluation based only on DNA concentration and purity. The framework successfully integrates biological, technical, economic, and environmental indicators, including reproducibility, scalability, cost efficiency, chemical safety, and sustainability, enabling a more reliable comparison of DNA extraction technologies for cross-kingdom genomic applications.

The study further concludes that artificial intelligence significantly improves DNA quality prediction by identifying complex relationships between extraction parameters and genomic outcomes. Machine learning models, particularly ensemble-based approaches such as Random Forest and XGBoost, demonstrated strong capability in predicting DNA yield, purity, integrity, and downstream molecular performance using laboratory and biological variables. The predictive performance was validated using statistical indicators, including accuracy, precision, recall, F1-score, MAE, RMSE, and R^2.

Additionally, the integration of explainable artificial intelligence enhances the transparency and practical usability of the proposed framework by identifying the most influential factors affecting DNA extraction quality. This capability supports evidence-based optimization of extraction conditions, reduces experimental failures, and improves reproducibility in genomic laboratories.

Overall, this research provides a novel pathway toward intelligent, sustainable, and standardized genomic workflows by integrating molecular biotechnology, machine learning, and explainable AI. The proposed framework has potential applications in microbial genomics, crop improvement, veterinary genetics, human molecular diagnostics, precision medicine, and environmental genomics. Future implementation across multiple laboratories and larger biological datasets will further strengthen its scalability and real-world applicability.

6.1 Contributions to Knowledge

This study makes significant contributions to knowledge by introducing an integrated AI-Driven Universal DNA Extraction and Quality Prediction Framework for Cross-Kingdom Genomic Applications that combines molecular biotechnology, sustainable extraction approaches, and artificial intelligence-based predictive analytics. The major contributions of this research are summarized as follows.

First, the study contributes a universal DNA extraction framework capable of supporting genomic analysis across multiple biological kingdoms, including bacteria, plants, animals, and humans. Unlike conventional extraction approaches that are generally optimized for specific sample categories, the proposed framework provides a standardized and adaptable workflow that improves DNA extraction consistency across diverse biological systems.

Second, the research contributes a comprehensive DNA extraction evaluation model by expanding comparison parameters beyond traditional measurements such as DNA concentration and purity. The framework introduces multidimensional evaluation criteria, including DNA yield, A260/A280 ratio, A260/A230 ratio, DNA integrity, PCR amplification success, sequencing compatibility, extraction time, reagent consumption, operational cost, reproducibility, scalability, chemical safety, and environmental sustainability. This provides a more complete approach for assessing genomic extraction technologies.

Third, the study contributes an AI-based DNA quality prediction approach that predicts genomic DNA performance using extraction-related variables. The integration of machine learning models, including Random Forest, XGBoost, Support Vector Machine, Artificial Neural Network, and ensemble learning approaches, demonstrates the capability of AI to identify complex relationships between biological characteristics, extraction conditions, and DNA quality outcomes.

Fourth, this research contributes to the field of explainable artificial intelligence in biotechnology by applying interpretable AI techniques to identify the most influential factors affecting DNA extraction performance. The identification of critical variables, such as sample type, lysis efficiency, reagent concentration, purification conditions, and incubation parameters, improves transparency and supports evidence-based laboratory optimization.

Fifth, the study contributes a sustainable genomic laboratory framework by reducing dependence on hazardous chemicals, minimizing reagent consumption, lowering operational costs, and improving workflow efficiency. The proposed framework supports environmentally responsible genomic research while maintaining high DNA quality and reliability.

Sixth, the research contributes new knowledge in cross-kingdom genomic standardization by demonstrating that a single AI-supported framework can adapt to different biological challenges associated with microbial, plant, animal, and human genomic samples. This advancement supports broader applications in microbial surveillance, agricultural improvement, veterinary genomics, clinical diagnostics, precision medicine, and environmental DNA analysis.

Finally, the study contributes a foundation for future intelligent genomic systems by demonstrating how artificial intelligence can transform traditional laboratory procedures into predictive, automated, and adaptive workflows. The proposed framework establishes a pathway toward next-generation genomic platforms where DNA extraction quality can be optimized dynamically using data-driven decision-making.

6.2 Limitations of the Study

Although the proposed AI-Driven Universal DNA Extraction and Quality Prediction Framework for Cross-Kingdom Genomic Applications demonstrates improved DNA extraction and prediction performance, several limitations should be considered. First, the study evaluates selected representative bacterial, plant, animal, and human samples; therefore, the findings may not fully represent the genetic and biochemical diversity of all organisms within each biological kingdom.

Second, the performance of AI-based DNA quality prediction models depends on the size, quality, and diversity of experimental datasets used for training and validation. Limited biological variations and laboratory conditions may affect model generalization when applied to new sample types or different research environments.

Third, the study focuses mainly on DNA extraction performance and quality prediction parameters, including DNA yield, purity, integrity, PCR success, sequencing compatibility, cost, time, and sustainability. However, extensive validation in large-scale genomic applications, clinical environments, and industrial sequencing platforms was beyond the scope of this research.

Finally, the proposed framework requires further optimization for fully automated laboratory implementation, real-time AI integration, and broader cross-species validation. Future research should incorporate larger multi-institutional datasets, additional biological samples, and advanced AI techniques to improve scalability and universal applicability.

6.3 Recommendations for Future Research

Future research should expand the AI-Driven Universal DNA Extraction and Quality Prediction Framework for Cross-Kingdom Genomic Applications by including a wider range of bacterial, plant, animal, and human samples to improve model generalization and universal applicability. Large-scale multi-institutional genomic datasets should be developed to enhance AI training, validation, and prediction reliability under different laboratory environments.

Further studies should investigate advanced artificial intelligence approaches, including deep learning, federated learning, and explainable AI, to improve real-time DNA quality prediction and automated extraction optimization. Integration with robotic laboratory systems and automated genomic platforms is recommended to achieve fully intelligent DNA extraction workflows.

Future research should also evaluate the framework using additional comparison parameters, including long-read sequencing performance, whole-genome assembly quality, metagenomic accuracy, environmental impact assessment, and large-scale cost–benefit analysis. Extended validation in clinical, agricultural, environmental, and industrial genomic applications will further confirm the practical value of the framework.

Finally, future investigations should focus on developing globally standardized AI-driven DNA extraction protocols that support sustainable, reproducible, and high-quality genomic analysis across diverse biological systems.

References

[1] Chandrashekar, K., Niranjan, V., Vishal, A., & Setlur, A. S. (2024). Integration of artificial intelligence, machine learning and deep learning techniques in genomics: Review on computational perspectives for NGS analysis of DNA and RNA-Seq data. Current Genomics, 19(9), 825–844. https://doi.org/10.2174/0115748936284044240108074937

[2] Chandrashekar, K., Niranjan, V., Vishal, A., & Setlur, A. S. (2024). Integration of artificial intelligence, machine learning and deep learning techniques in genomics: Review on computational perspectives for NGS analysis of DNA and RNA-Seq data. Current Genomics, 19(9), 825–844.

[3] Cornet, L., & Baurain, D. (2022). Contamination detection in genomic data: More is not enough. Genome Biology, 23, 60. https://doi.org/10.1186/s13059-022-02619-9

[4] Council for International Organizations of Medical Sciences. (2021). International ethical guidelines for health-related research involving humans. CIOMS.

[5] Galla, G., Praeg, N., Rzehak, T., Sprecher, E., Colla, F., Seeber, J., Illmer, P., & Hauffe, H. C. (2024). Comparison of DNA extraction methods on different sample matrices within the same terrestrial ecosystem. Scientific Reports, 14, 8715. https://doi.org/10.1038/s41598-024-59086-4

[6] Green, M. R., & Sambrook, J. (2019). Molecular cloning: A laboratory manual (4th ed.). Cold Spring Harbor Laboratory Press.

[7] Integrating Artificial Intelligence in Next-Generation Sequencing: Advances, Challenges, and Future Directions. (2025). Current Issues in Molecular Biology, 47(6), 470. https://doi.org/10.3390/cimb47060470

[8] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255–260. https://doi.org/10.1126/science.aaa8415

[9] Kim, C., Pongpanich, M., & Porntaveetus, T. (2024). Unraveling metagenomics through long-read sequencing: A comprehensive review. Journal of Translational Medicine, 22, 111. https://doi.org/10.1186/S12967-024-04917-1

[10] Leung, M. K. K., Delong, A., Alipanahi, B., & Frey, B. J. (2016). Machine learning in genomic medicine: A review of computational approaches and applications. Genome Medicine, 8, Article 73. https://doi.org/10.1186/s13073-016-0338-7

[11] Li, P., et al. (2022). Solid-phase extraction methods for nucleic acid separation: A review. Journal of Separation Science, 45(1), 172–184. https://doi.org/10.1002/jssc.202100295

[12] Libbrecht, M. W., & Noble, W. S. (2015). Machine learning applications in genetics and genomics. Nature Reviews Genetics, 16, 321–332. https://doi.org/10.1038/nrg3920

[13] Maqsood, K., Hagras, H., & Zabet, N. R. (2024). An overview of artificial intelligence in the field of genomics. Discover Artificial Intelligence, 4(9). https://doi.org/10.1007/s44163-024-00103-w

[14] Montgomery, D. C. (2020). Design and analysis of experiments (10th ed.). John Wiley & Sons.

[15] National Academies of Sciences, Engineering, and Medicine. (2017). Human genome editing: Science, ethics, and governance. The National Academies Press. https://doi.org/10.17226/24623

[16] National Research Council. (2011). Guide for the care and use of laboratory animals (8th ed.). National Academies Press. https://doi.org/10.17226/12910

[17] O'Connor, O., & McVeigh, T. P. (2025). Increasing use of artificial intelligence in genomic medicine for cancer care: The promise and potential pitfalls. BJC Reports, 3, Article 20.

[18] Organisation for Economic Co-operation and Development. (2019). Recommendation of the Council on Artificial Intelligence. OECD Publishing.

[19] Osinski, B. L., BenTaieb, A., Ho, I., Jones, R. D., Joshi, R. P., Westley, A., Carlson, M., Willis, C., Schleicher, L., Mahon, B. M., et al. (2022). AI-augmented histopathologic review using image analysis to optimize DNA yield and tumor purity from FFPE slides. arXiv. https://arxiv.org/abs/2203.13948

[20] Pfeifer, G. P., & Jin, S.-G. (2024). Methods and applications of genome-wide profiling of DNA damage and rare mutations. Nature Reviews Genetics, 25, 846–863. https://doi.org/10.1038/s41576-024-00748-4

[21] Pinto, Y., & Bhatt, A. S. (2024). Sequencing-based analysis of microbiomes. Nature Reviews Genetics, 25, 829–845. https://doi.org/10.1038/s41576-024-00746-6

[22] Recent Advancements and Emerging Techniques in Nucleic Acid Isolation, Amplification, and Detection from Diverse Complex Matrices of Human Interest. (2025). Trends in Analytical Chemistry, 185, 118172. https://doi.org/10.1016/j.trac.2025.118172

[23] Sambrook, J., & Russell, D. W. (2001). Molecular cloning: A laboratory manual (3rd ed.). Cold Spring Harbor Laboratory Press.

[24] Sheershika, & Ram, M. (2024). Advances in DNA extraction techniques: A comprehensive review of methods and applications. Plant Cell Biotechnology and Molecular Biology, 25(5–6), 30–42. https://doi.org/10.56557/pcbmb/2024/v25i5-68683

[25] Stammnitz, M. R., Hartman Scholz, A., & Duffy, D. J. (2024). Environmental DNA without borders. EMBO Reports, 25, e59624. https://doi.org/10.1038/s44319-024-00264-w

[26] Taddese, A. A., Addis, A. C., & Tam, B. T. (2025). Data stewardship and curation practices in AI-based genomics and automated microscopy image analysis for high-throughput screening studies: Promoting robust and ethical AI applications. Human Genomics, 19, 16. https://doi.org/10.1186/s40246-025-00716-x

[27] Tan, S. C., & Yiap, B. C. (2009). DNA, RNA, and protein extraction: The past and the present. Journal of Biomedicine and Biotechnology, 2009, 574398. https://doi.org/10.1155/2009/574398

[28] Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

[29] Tsui, W. H. A., Ding, S. C., Jiang, P., & Lo, Y. M. D. (2025). Artificial intelligence and machine learning in cell-free-DNA-based diagnostics. Genome Research, 35(1), 1–19. https://doi.org/10.1101/gr.278413.123

[30] UNESCO. (2021). UNESCO science report: The race against time for smarter development. UNESCO Publishing.

[31] van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. Trends in Genetics, 30(9), 418–426. https://doi.org/10.1016/j.tig.2014.07.001

[32] World Health Organization. (2021). Ethics and governance of artificial intelligence for health. World Health Organization.