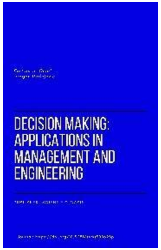
Decision Making: Applications in Management and Engineering, Volume 8, Issue 1 (2025) 401-418

SCIENTIFIC OASIS

Decision Making: Applications in
Management and Engineering

Journal homepage: www.dmame-journal.org
ISSN: 2560-6018, eISSN: 2620-0104



DECISION MAKING: APPLICATIONS IN MANAGEMENT AND ENGINEERING

Role of Artificial Intelligence in Bioinformatics and Information Extraction Systems

Maram Y. Al-Safarini[1*] Rania A. Baashirah[2]

1 Computer Science Department, Zarqa University, Jordan, Email: maram@zu.edu.jo
2 Software Engineering Department, University of Business & Technology. Email: r.baashirah@ubt.edu.sa

ARTICLE INFO

Article history:
Received 5 July 2024
Received in revised 14 October 2024
Accepted 25 March 2025
Available online 30 March 2025

Keywords:
Artificial Intelligence, Information Extraction Systems, Bioinformatics.

ABSTRACT

Artificial Intelligence (AI) is transforming numerous domains, including bioinformatics and information extraction systems, by advancing data processing capabilities, enhancing precision, and facilitating automation. The primary aim of this research is to examine the function of AI within the realms of bioinformatics and information extraction through a combination of quantitative and qualitative approaches. The quantitative component involved a sample of 152 participants, recognised as experts in bioinformatics and data science. The survey findings underscore the prevalent integration of AI in bioinformatics, particularly through the utilisation of reinforcement learning, neural networks, and natural language processing (NLP). Furthermore, AI substantially improves the analysis of biological data; however, it encounters challenges such as limited model interpretability and the lack of data standardisation. Notwithstanding these obstacles, AI-driven innovations in disease prognosis, personalised healthcare, and pharmaceutical development are anticipated to shape the future trajectory of bioinformatics. The qualitative findings, derived through thematic analysis encompassing core themes and sub-themes, reveal that AI significantly contributes to accelerating and refining information extraction processes. Technologies such as machine learning, NLP, and neural networks are instrumental in enhancing data processing efficiency. Nonetheless, issues such as inadequate data quality, elevated computational expenses, and the intricacy of AI models remain persistent. Looking ahead, AI is projected to integrate more effectively with big data infrastructures, enable real-time information extraction, and deliver increasingly tailored solutions. The study concludes with several policy recommendations to guide future implementation strategies.

1. Introduction

The ideology underpinning bioinformatics constitutes an interdisciplinary fusion of molecular biology, genetics, computer science, mathematics, and statistics, employed to examine biological data through a suite of computational methodologies and models [1; 2]. The principal objective of bioinformatics is to address intricate biological issues, particularly those involving voluminous

---

[*] *Corresponding author.*
E-mail address: maram@zu.edu.jo

https://doi.org/10.31181/dmame8120251376

datasets. Consequently, its core emphasis lies in deciphering biological mechanisms at the molecular level and deriving meaningful interpretations from the data acquired [3]. Among the foundational strategies in bioinformatics is the statistical analysis of biological information. This approach facilitates the construction of computational models aimed at resolving scientific problems, alongside the continual refinement and validation of analytical algorithms [4]. Such advantages constitute the remarkable contributions of bioinformatics to biological research. In this regard, bioinformatics may be likened to contemporary data analysis practices, where previous constraints stemmed from limited access to information [5]. However, advancements in DNA sequencing technologies have significantly alleviated this barrier. The prevailing challenge now lies in the interpretation of extensive datasets—comparable to analysing web traffic—which necessitates the utilisation of computer-assisted tools. Moreover, the decoding of the human genome increasingly depends upon computational techniques, as bioinformatics seeks to derive insightful and actionable outcomes from complex biological data.

Computational biology, in particular, is oriented towards achieving substantive advancements within the broader domain of bioinformatics (BI), especially in healthcare and the analysis of biological information [6]. The application of AI, inspired by the problem-solving mechanisms observed in natural systems, continues to evolve and augment proficiency in this field [7]. This mixed-methods study investigates the contribution of AI to bioinformatics and information extraction systems, emphasising how AI-based approaches enhance data interpretation, pattern recognition, and predictive modelling within biological sciences. The study provides a comprehensive overview of sophisticated computational techniques that facilitate the efficient extraction and understanding of complex biological information. It also explores pivotal AI resources and tools, as documented in academic literature and digital repositories, rendering the findings valuable for practitioners engaging with large-scale bioinformatics data.

It is widely acknowledged that a singular methodology cannot resolve all biological challenges due to population-specific genetic variations and mutational dynamics. In the present context, global transformations are increasingly shaped by AI, which constitutes a cornerstone of the Fourth Industrial Revolution [8-10]. Consequently, the development of bioinformatics solutions has become imperative. In this pursuit, AI demonstrates exceptional potential to yield accurate insights and stimulate innovation, thereby facilitating progress in bioinformatics and enabling data-informed decision-making with beneficial societal implications. This is evident in AI-driven bioinformatics activities, which include accurate data analysis, expedited information processing, improved operational efficiency, and minimised error rates in data-centric systems [11]. Moreover, as the global elderly population expands due to medical and technological progress, bioinformatics also plays a crucial role in addressing the health-related complexities associated with ageing.

Over the past three decades, computational biology has evolved into a well-established scientific discipline, making substantial contributions to our comprehension of diverse biological phenomena. Nonetheless, researchers in this domain, across various institutions, continue to encounter professional development challenges. Computational biology relies significantly on interdisciplinary collaboration and mutual appreciation among scientists from diverse fields and geographical regions [12]. Thus, it is essential for researchers to acknowledge and respect each other's specialisations and motivations. Scientists should not adopt a siloed approach or regard any field as subordinate. For instance, computational biologists wish to avoid being perceived merely as individuals who "run the numbers," just as experimental biologists object to being characterised solely as "data generators." Hence, computational biology encompasses a wide array of domains, including statistics, database architecture, clinical informatics, genetics, epigenetics, genomics, proteomics, biomedical imaging, single-cell analysis, structural prediction, algorithmic design, and machine learning, among others.

Recent innovations in high-speed processing technologies have led to the generation of vast biological datasets, ranging from genetic sequences to complex multi-omics data. While these datasets offer invaluable insights into life processes, their scale and complexity often exceed the capabilities of traditional analytical methods. Computational biology, leveraging mathematics, statistics, and algorithmic computation, has been pivotal in resolving such complex biological inquiries [13]. However, as the intricacy of these problems escalates, AI—particularly machine learning and deep learning—has emerged as an indispensable tool. The rationale lies in AI's capacity to process and interpret extensive and multifaceted datasets with exceptional precision, catalysing a paradigm shift in computational biology. Through the integration of AI with computational biology, researchers are uncovering novel insights in areas such as genomics, drug discovery, and personalised medicine, thereby propelling considerable progress in healthcare innovation [23].

Information Extraction (IE) has gained prominence as an expanding discipline, bolstered by the widespread proliferation of social media platforms that connect millions globally while generating vast quantities of textual content. Extracted data is deployed in numerous domains, ranging from targeted marketing strategies to intelligence-gathering for defence and security operations. AI encompasses diverse techniques such as machine learning, logical inference, information retrieval, computational linguistics, and pattern recognition. IE is especially effective in retrieving pertinent information from data that is unstructured or semi-structured in nature. In the current era of data deluge, IE has become increasingly complex. Large, multidimensional unstructured datasets—arising from big data—pose significant analytical challenges. However, the amalgamation of computational power with AI technologies is proving transformative in managing such datasets. These innovations offer superior analytical capabilities compared to traditional IE methods. At present, neural and adaptive-based computing models demonstrate considerable promise in enhancing information extraction efficacy.

2. Literature Review

The application of AI within the domain of bioinformatics has been evident for over two decades. AI has become an integral component of bioinformatics research [14]. They further highlighted that numerous AI-linked algorithms have now become readily accessible, enabling researchers to utilise pre-existing AI systems for the analysis and classification of biological data. Nevertheless, the abundance of intelligent methodologies available presents challenges in selecting the most appropriate technique for a specific dataset, particularly within the field of bioinformatics. To address this complexity, researchers require tools capable of not only organising data effectively but also presenting it in a clear, meaningful manner, complemented by explanatory notes, accurate estimates, and contextual relevance.

This article explores the utilisation of AI in bioinformatics and computational molecular biology, particularly within DNA sequencing. These fields have evolved in response to the increasing necessity for biologists to interpret extensive genomic datasets. A significant driver of the heightened attention towards bioinformatics and DNA sequencing is the complexity involved in studying organismal evolution, especially when confronting incomplete or flawed data. Furthermore, the article examines various software tools developed to support researchers, encompassing those designed for searching, categorising, and analysing biological databases, as well as simulating biological experiments. Consequently, AI-driven software enables scientists to detect diverse patterns, refine analyses, and gain profound insights into biological data.

The study conducted by Nicolas [15] underscores the longstanding relationship between AI and bioinformatics. The author illustrates that bioinformatics has historically posed intricate challenges well-suited to AI interventions. In his discussion, Nicolas [15] introduces several critical issues to demonstrate the dynamic character of bioinformatics, providing an accessible entry point for emerging researchers. Alongside identifying key questions, the author also reviews notable advancements within the discipline. The primary objective of the study is to facilitate improved access for bioinformaticians, bioanalysts, and biologists to the latest methodologies and innovations. Additionally, a recent mini-review by Jamialahmadi et al. [16] elaborates on how the integration of AI has revolutionised the understanding and analysis of biological data. Adhering to PRISMA guidelines, the authors conducted a systematic review of relevant studies published between 2018 and 2024 across databases such as PubMed, Embase, and Google Scholar. The review focuses on the deployment of machine learning, deep learning, and NLP within bioinformatics, addressing complex areas such as genome sequencing, protein structure prediction, drug discovery, systems biology, personalised medicine, imaging, signal processing, and text mining. This study captures the rapidly evolving landscape of AI-enabled tools and algorithms, asserting that AI not only accelerates research but also transforms data interpretation, pushing the boundaries of innovation within biomedical sciences. As AI progresses, it continues to resolve previously insurmountable challenges, facilitating faster and more accurate biological discoveries.

Moreover, the management of extensive datasets generated through high-throughput biotechnologies demands efficient data handling and sophisticated interpretative methods [17]. Consequently, computational intelligence and machine learning techniques have become vital for analysing biomolecular data and for constructing models that encapsulate biological knowledge. One of the earliest applications of machine learning in bioinformatics dates to the 1980s, with the use of the perception, inspired by neural cell functions, to identify translation initiation sites in prokaryotes. Since then, an array of computational intelligence and machine learning methods has been developed and refined to address a wide range of bioinformatics challenges [17].

Contemporary studies have also emphasised the critical role of AI and machine learning in bioinformatics. For instance, research by Shree et al. [18] explores the complexities of medical classification and analysis, highlighting their pivotal role in assisting healthcare professionals with diagnostic and therapeutic decisions. Various medical imaging modalities—including X-rays, ultrasounds, MRIs, and histopathological images, are increasingly analysed through machine learning and AI-supported systems. These systems employ advanced algorithms to interpret images, offering supplementary insights by detecting anomalies, classifying diseases, and evaluating treatment efficacy. They act as strategic aids, providing healthcare practitioners with detailed measurements, visual indicators, and probability estimates, thereby enhancing diagnostic accuracy and patient management. The early detection and appropriate treatment of diseases are critical, as they substantially improve patient outcomes and can be lifesaving. Thus, the overarching aim of such research is to advance the precision and reliability of machine learning applications in medical diagnostics, ultimately contributing to better healthcare delivery.

3. Methods and Materials

This study employed a mixed-methods approach to investigate the role of AI within the fields of bioinformatics and information extraction systems. Both quantitative and qualitative methods were utilised, reflecting their methodological significance [19-21]. Accordingly, data were collected and analysed through both approaches. For the quantitative aspect, frequency distribution analysis was applied to a sample dataset, assisting in the identification of patterns, trends, and distributions concerning AI applications in bioinformatics, associated concepts, prevailing challenges, and prospective opportunities. This facilitated a clearer understanding of the prevalence and significance of AI applications within the domain. The findings of this analysis provided a fundamental basis for comprehending the influence of AI technologies in bioinformatics and information extraction systems. Data for the quantitative analysis was collected through a questionnaire, detailed in subsections 4.1 to 4.7. The questionnaire was distributed among a range of professionals, including bioinformaticians, data scientists, computational biologists, researchers, healthcare professionals, and other related roles. Initially, 200 questionnaires were disseminated, of which 167 were returned. Following verification, 15 questionnaires were deemed invalid due to incorrect or incomplete responses, resulting in a final size sample of 152 used for the frequency distribution analysis.

For the qualitative component, thematic analysis was performed on semi-structured interviews conducted with 12 participants, comprising researchers, bioinformaticians, and AI specialists. The semi-structured format allowed for in-depth and flexible responses. Thematic analysis facilitated the identification and examination of patterns and themes emerging from the interview data. Various themes and sub-themes were developed and discussed concerning AI and information extraction systems. By integrating frequency distribution analysis with thematic analysis, the study effectively balanced quantitative breadth with qualitative depth. The frequency analysis offered a comprehensive overview of the data, whilst the thematic analysis provided nuanced, human-centred insights.

4. Results and discussions

4.1 Demographics

The characteristics of the study's respondents are comprehensively presented in Table 1, detailing their age, current professional role, and years of experience in the field of bioinformatics.

Table 1
Demographics

Age	Frequency	%
20-30	61	40.13158
31-40	53	34.86842
41-50	16	10.52632
>50	22	14.47368
Total	152	100
1. What is your current role? (Select one)	Frequency	%
- Bioinformatician	31	20.39474
- Data Scientist	55	36.18421
- Computational Biologist	19	12.5
- Researcher	27	17.76316
- Healthcare Professional	11	7.236842
- Other	9	5.921053
Total	152	100
. How many years of experience do you have in bioinformatics or related fields? (Select one)	Frequency	%
- Less than 2 Years	51	33.55263
- 2-5 Years	46	30.26316
- 5-10 Years	32	21.05263
- More than 10 Years	23	15.13158
Total	152	100

The age distribution indicates that most participants fall within the 20–30 years age bracket. The second largest group, comprising 35% of respondents, falls within the 31–40 years range. Conversely, only 10% and 14% of participants are aged between 41–50 years and over 50 years, respectively. The distribution of respondents' roles demonstrates a diverse professional representation. The largest group consists of Data Scientists, accounting for 36% of the sample. This is followed by Bioinformaticians, representing 20% of respondents. Researchers constitute a further 18%, while

Computational Biologists make up 13% of the participants. The remaining respondents are Healthcare Professionals (7%) and individuals from other professional backgrounds (6%).

Regarding years of experience in bioinformatics, the data reveal a well-balanced range. A significant proportion of respondents, 34%, have less than two years of experience, indicating a substantial presence of relatively new entrants to the field. Meanwhile, 30% of participants possess between two to five years of experience. Furthermore, 21% have between five to ten years of experience, whilst 15% have accumulated over ten years of expertise, reflecting a combination of both emerging and seasoned professionals. Overall, the survey highlights that most participants are relatively young, with 40% aged between 20–30 years, and that most have limited experience in the field, although a notable number possess considerable expertise. The distribution of these properties is visually depicted in Figure 1.

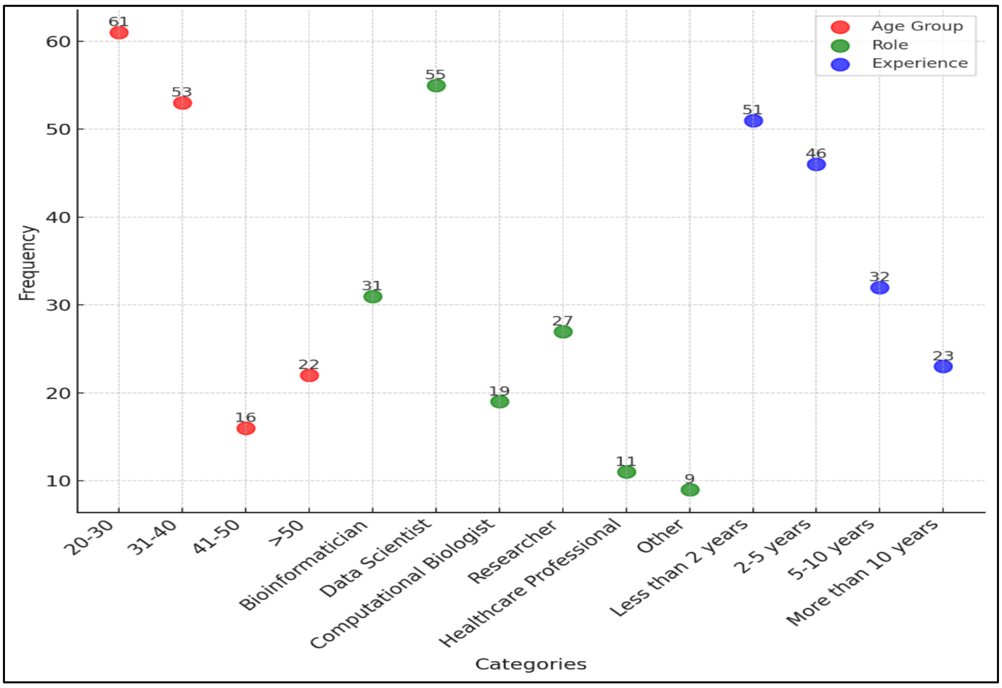


Fig.1. Demographics

4.2 AI Tools and Techniques in Bioinformatics

This section examines the AI tools and techniques associated with bioinformatics. The frequency distribution analysis provides several valuable insights. For instance, as presented in Table 2, the question “How familiar are you with AI-based bioinformatics tools? (Select one)” reveals that 61.84% of respondents—equating to 94 out of 152—reported being very familiar with such tools and techniques. In contrast, only 26 participants indicated a lack of familiarity with these tools. Additionally, 21% of the sample noted that they were somewhat familiar with the technologies in question. Further data in Table 2, related to the question “What is your current role?”, indicate that the largest proportion of respondents are Data Scientists, comprising 36.18% of the total sample, or 55 out of 152 participants. This is followed by Bioinformaticians, who account for 20.39% of the sample, representing 31 individuals. Researchers constitute 17.76% of the total, or 27 participants. A smaller share is represented by Computational Biologists, comprising 12.5% of the sample. Meanwhile, Healthcare Professionals make up 7.24% of respondents, and 5.92% fall under the 'Other' category. Additionally, the distribution of respondents by years of experience in bioinformatics, as shown in Table 2, indicates that 33.55%—equivalent to 51 participants—have less than two years of experience.

Table 2
AI Tools and Techniques in Bioinformatics

How familiar are you with AI-based bioinformatics tools? (Select one)	Frequency	%
- Not Familiar	26	17.11
- Somewhat Familiar	32	21.05
- Very Familiar	94	61.84
Total	152	100
Which AI techniques do you apply in your bioinformatics work? (Select all that apply)	Frequency	--
- Machine Learning [22]	62	--
- Deep Learning (DL)	33	--
- Natural Language Processing (NLP)	85	--
- Neural Networks	105	--
- Reinforcement Learning	167	--
- Other	51	--
What bioinformatics tools powered by AI do you use regularly? (Select all that apply)	Frequency	--
- AlphaFold	32	--
- BLAST (with AI features)	66	--
- GATK (Genome Analysis Toolkit)	19	--
- DeepGene	78	--
- CATH (AI-based protein structure prediction)	108	--
- Other	23	--

A further 30.26%, or 46 respondents, reported having between two and five years of experience. Approximately 21.05% of the sample, or 32 participants, have five to ten years of experience. Lastly, 15.13% of respondents indicated having more than ten years of professional experience in the field.

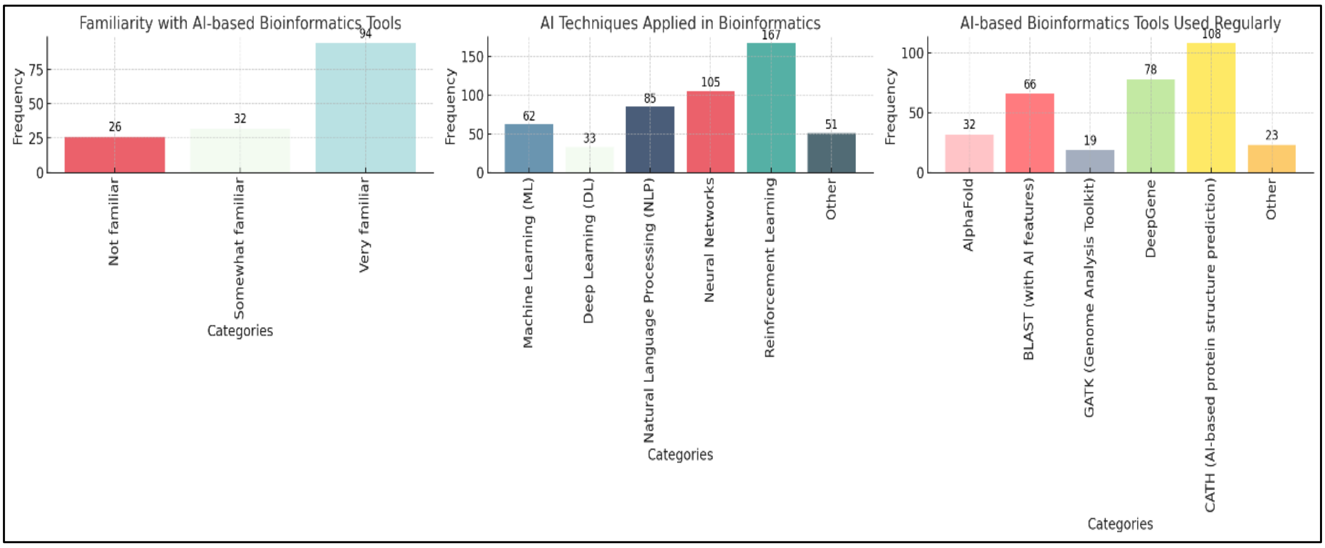


Fig.2. AI Tools and Techniques in Bioinformatics

4.3 AI Integration in Data Analysis

The integration of AI into bioinformatics represents another key area explored within this study. Table 3 presents a comprehensive distribution across various classifications of the relevant queries. Regarding the question, “How often do you apply AI or machine learning in bioinformatics data analysis? (Select one)”, the distribution of responses is as follows: 0, 27, 12, 32, and 81 participants selected the options 'never', 'rarely', 'occasionally', 'frequently', and 'always', respectively. Expressed as percentages, these categories correspond to 0.00% (lowest), 17.76%, 7.89%, 21.05%, and 53.29% (highest), respectively. These results reveal that over 53% of respondents consistently utilise AI or machine learning in their bioinformatics data analyses. Specifically, the majority, 53.29%, reported always employing AI or machine learning, followed by 21.05% who use these methods frequently. Conversely, 17.76% of participants indicated rare use, while 7.89% noted occasional use. Notably, no respondents reported that they never incorporate AI or machine learning into their work.

Table 3
AI Integration in Data Analysis

How often do you apply AI or machine learning in bioinformatics data analysis? (Select one)		
- Never	0	0
- Rarely	27	17.76316
- Occasionally	12	7.894737
- Frequently	32	21.05263
- Always	81	53.28947
Total	152	100
For which types of data analysis do you use AI in bioinformatics? (Select all that apply)		
- Sequence Alignment and Analysis	32	
- Protein Structure Prediction	62	
- Gene Expression Analysis	127	
- Data Clustering and Pattern Recognition	62	
- Biomarker Identification	106	
- Disease Prediction and Diagnosis	52	
How effective do you find AI-based tools in bioinformatics data analysis? (Select one)	Frequency	%
- Very Ineffective	4	2.631579
- Somewhat Ineffective	64	42.10526
- Neutral	19	12.5
- Somewhat Effective	31	20.39474
- Very Effective	34	22.36842
Total	152	100

In terms of specific applications, Gene Expression Analysis emerged as the most frequently cited area benefiting from AI, receiving 127 responses. This was followed by Biomarker Identification with 106 responses, and Protein Structure Prediction with 62 responses. Furthermore, when respondents were asked to evaluate the effectiveness of AI tools in bioinformatics data analysis, 42.11% of the sample considered the tools somewhat ineffective. In contrast, 20.39% perceived them as somewhat effective. A smaller proportion, 22.37%, found the tools to be very effective, whilst only 2.63% classified them as very ineffective. These findings suggest that although many respondents find AI tools beneficial, there remains considerable scope for enhancing their effectiveness. The integration of AI in data analysis is further illustrated graphically in Figure 3.

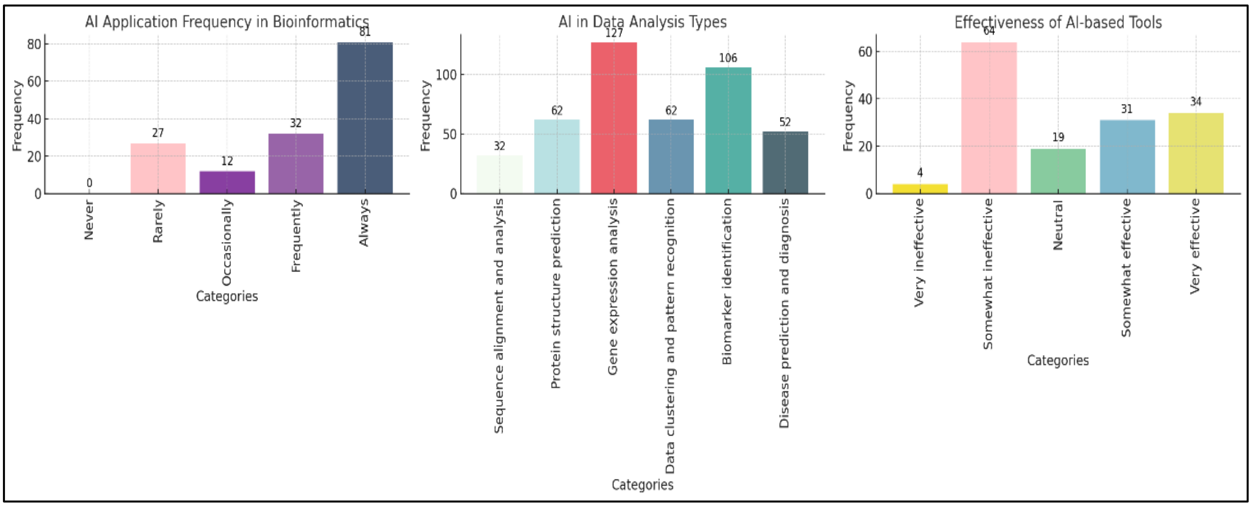


Fig.3. AI Integration in Data Analysis

4.4 Information Extraction in Bioinformatics

This section addresses the utilisation of information extraction techniques in bioinformatics, based on the respondents' views. The initial query, “Do you use information extraction techniques to handle biological data? (Select one)”, revealed that 141 respondents affirmed their use, while only 11 indicated otherwise. This distribution provides substantial support for the notion that information extraction techniques are widely employed by the respondents in handling biological data. The subsequent query, “Which information extraction techniques do you use in bioinformatics?”, offers further insights. The findings show that Image Analysis, particularly in relation to microscopy or imaging data, is the most frequently utilised technique, cited by 133 respondents. This is followed by Natural Language Processing, employed by 62 respondents. Pattern Recognition also appears to be widely applied, with 64 participants indicating its use for extracting meaningful information from biological datasets.

Table 4
Information Extraction in Bioinformatics

Do you use information extraction techniques to handle biological data? (Select one)	Frequency
- Yes	141
- No	11
Total	
Which information extraction techniques do you use in bioinformatics? (Select all that apply)	Frequency
- Text Mining	12
- Natural Language Processing (NLP)	62
- Data Mining	19
- Image Analysis (for microscopy or imaging data)	133
- Pattern Recognition	64
- Other	88
How effective do you find AI-based information extraction tools in bioinformatics? (Select one)	
- Very Ineffective	5
- Somewhat Ineffective	17
- Neutral	2
- Somewhat Effective	49
- Very Effective	79
Total	152

Interestingly, Text Mining is reported by only 12 respondents, suggesting it is comparatively less common. In addition, Data Mining is utilised by 19 respondents, while 88 individuals reported employing other techniques not specified among the listed options. This diversity in methods highlights the broad range of approaches adopted by bioinformaticians and researchers for extracting information from complex biological datasets. Finally, the query, “How effective do you find AI-based information extraction tools in bioinformatics?”, revealed that 79 respondents (the majority) considered these tools to be very effective, while 49 viewed them as somewhat effective. A smaller proportion, 17 respondents, regarded the tools as somewhat ineffective, and only 5 respondents found them very ineffective. These findings suggest a generally positive perception of the effectiveness of AI-based tools for information extraction in bioinformatics, although there remains a minority who perceive a need for further improvements. The distribution of these responses is also illustrated graphically in Figure 4. In addition, Information extraction in Bioinformatics is reported in Table 4.

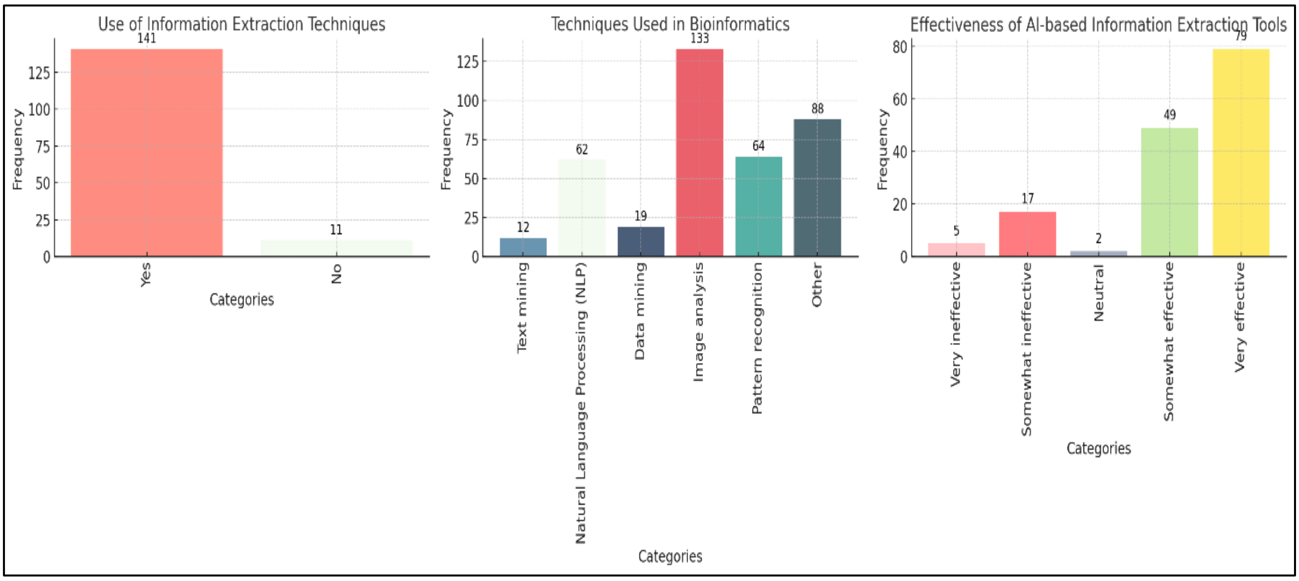


Fig.4. Information Extraction in Bioinformatics

4.5 Major Challenges While Using Information Extraction Techniques in Bioinformatics

The study also explored the challenges associated with the use of information extraction techniques in bioinformatics, with respondent feedback presented in Table 5. The frequency distributions indicate that multiple categories were selected by participants, highlighting various ongoing issues in the field. One of the key concerns identified was difficulty handling unstructured data, such as free-text records, which was cited by 89 respondents, placing it in the sixth category of Table 5. This underscores a persistent issue in bioinformatics, where the processing of non-standardised and poorly organised data remains complex.

The second major challenge relates to model interpretation and explainability, referenced by 71 respondents. This suggests that despite the analytical power of AI and machine learning models, their inherent complexity often poses difficulties for researchers in interpreting and clearly explaining the outcomes they produce. Lack of standardisation in tools or techniques was noted by 51 participants, pointing to complications caused by inconsistencies in the methodologies and tools used across studies. This lack of uniformity undermines reproducibility and comparability in bioinformatics research. Additionally, data quality issues, such as noise or missing data, were reported by 24 respondents. These problems significantly affect the reliability of outcomes, as inaccurate or incomplete datasets can distort research findings, particularly when handling large-scale biological data.

Table 5
Major Challenges While Using Information Extraction Techniques in Bioinformatics

What are the major challenges you face when using information extraction techniques in bioinformatics? (Select all that apply)	Frequency
Data Quality Issues (e.g., noisy or missing data)	24
Lack of Standardization in Tools or Techniques	51
Computational Power Limitations	11
Difficulty in Model Interpretation and Explain Ability	71
Lack of Domain-Specific Knowledge Integration	19
Difficulty Handling Unstructured Data (e.g., free-text records)	89
Insufficient Training Data for AI Models	27
Integration of Diverse Data Types (e.g., genomic, clinical, environmental data)	37

Insufficient training data for AI models was acknowledged by 27 respondents (positioned seventh in Table 5), highlighting the limitations in data availability that hinder the capacity of these models to produce precise predictions or classifications. Computational power limitations were reported by 11 respondents, indicating the demanding processing requirements associated with bioinformatics analyses, which can exceed the available computational resources. Furthermore, lack of domain-specific knowledge integration was cited by 19 respondents, illustrating a gap between technical AI applications and the specialised knowledge necessary for accurate bioinformatic interpretations. Lastly, the integration of diverse data types, such as genomic, clinical, and environmental information, was raised as a challenge by 37 participants. The complexity of aligning these varied data sources poses significant barriers to effective and comprehensive bioinformatic analysis.

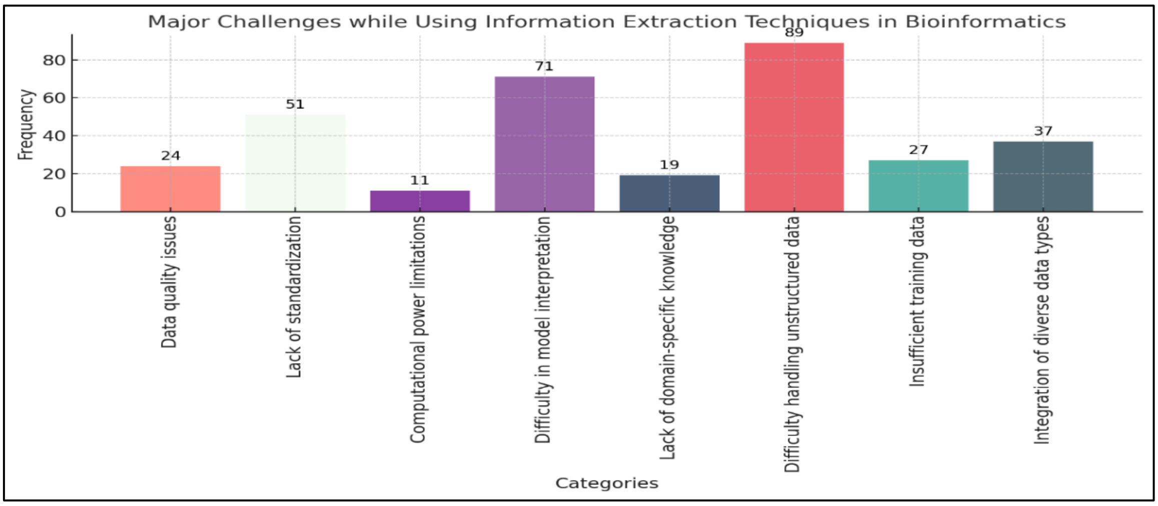


Fig.5. Major Challenges While Using Information Extraction Techniques in Bioinformatic

4.6 AI Tools and Techniques in Computational Biology and AI Integration in Data Analysis

The present section investigates respondents' perspectives on the application of AI tools and techniques in computational biology. The findings reveal that 35 respondents highlighted the use of AI in biological simulations, while 61 mentioned its application in evolutionary modelling. Additionally, 22 respondents referred to metabolic network analysis, and 84 discussed gene regulatory network modelling. A smaller group, 18 respondents, indicated that other AI-based techniques are also employed in this field. When evaluating the role of AI in computational biology, 33 respondents acknowledged its contribution to enhancing protein structure predictions, while 67 participants noted its utility in simulating biological processes.

Table 6
AI and Computational Biology

Which AI techniques do you apply in computational biology research? (Select all that apply)	Frequency
AI-Driven Biological Simulations	35
Evolutionary Modelling	61
Metabolic Network Analysis	22
Gene Regulatory Network Modelling	84
Other	18
AI Integration in Data Analysis	
How does AI contribute to computational biology in your work? (Select all that apply)	Frequency
Enhancing Protein Structure Predictions	33
Simulating Biological Processes (e.g., metabolic pathways)	67
Modelling Gene Regulatory Networks	53
AI-Driven Drug Design and Molecular Interaction Studies	88

Other	24
Information Extraction in Bioinformatics and Computational Biology	
12A. How do you use AI for extracting biological insights in computational biology? (Select all that apply)	Frequency
AI-Driven Text Mining for Biological Hypothesis Generation	61
Extracting Patterns from Large-Scale Biological Datasets	98
Automated Extraction of Relationships in Gene Expression Data	107
Other	58

Gene regulatory network modelling was also cited as significant by 53 respondents. Notably, the most frequently mentioned contribution, reported by 88 respondents, was AI-driven drug design and molecular interaction studies, underscoring the growing interest in AI's potential within drug discovery and molecular research. Furthermore, an additional 24 responses highlighted other ways AI contributes to data analysis in computational biology. This distribution is also presented in Figure 6. In addition, AI and computational Biology are presented in Table 6.

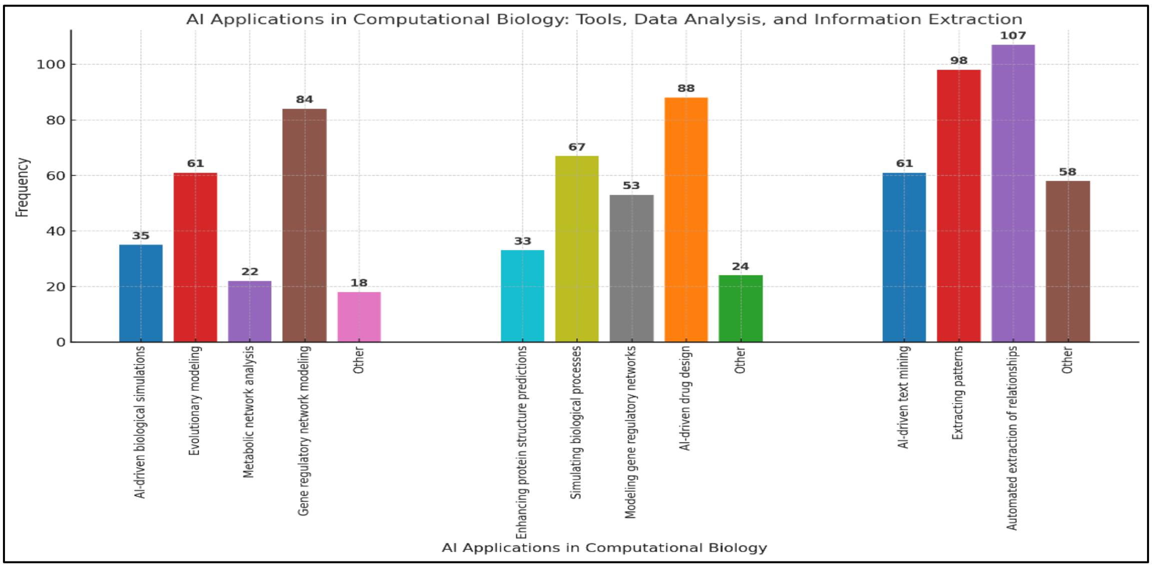


Fig.6. AI and Computational Biology

4.7 Future Trends and Opportunities

The final quantitative investigation delves into future trends and opportunities as perceived by the respondents. When asked, "How do you foresee the role of AI in bioinformatics in the next 5-10 years?", the responses indicate significant optimism. Approximately 53% of respondents foresee substantial growth in AI's role within this field, while 14% predict a moderate level of growth. However, only 12.50% anticipate a decline in AI's role over the next 5-10 years. A minority of 11 respondents expressed uncertainty regarding this future trajectory. In terms of AI’s potential contribution to personalised medicine through bioinformatics, the responses reflect a generally positive outlook. Specifically, 30.26% of participants believe AI will make a very high contribution, particularly in providing tailored treatments based on genetic, environmental, and lifestyle factors. A slightly larger group, 36.18%, anticipates a moderate contribution, while 5.26% foresee a low level of AI involvement. This lower expectation may stem from scepticism regarding AI's capacity to manage the complexity of personalised medicine. Furthermore, 15.13% of respondents feel that AI will make no contribution at all, potentially reflecting doubts about its practical applicability in this domain. Lastly, 13.16% of respondents were unsure, suggesting a lack of clarity about how AI will influence personalised medicine.

Regarding potential breakthroughs in AI and bioinformatics, the responses highlighted various possibilities. The most frequently cited breakthrough, selected by 36.18% of respondents, is disease prediction and diagnosis, which reflects a strong belief in AI's ability to enhance the accuracy and efficiency of disease identification. Personalised medicine ranked second, with 21.71% of responses, indicating widespread expectations for AI to aid in the customization of treatments for individual patients. Additionally, 15.13% of respondents identified genomic sequencing and analysis as another promising area, while 12.50% viewed drug discovery as a key breakthrough where AI could significantly accelerate the development of new medications. Finally, 14.47% of respondents selected "other", suggesting alternative or less conventional ideas about the potential future impact of AI in bioinformatics. This distribution is also illustrated in Figure 7. The future trends and opportunities mentioned in Table 7.

Table 7
Future Trends and Opportunities

13. How do you foresee the role of AI in bioinformatics in the next 5-10 years? (Select one)	Frequency	%
- Will Significantly Grow	81	53.29
- Will Grow Moderately	22	14.47
- Will Remain the Same	19	12.50
- Will Decrease	19	12.50
- Unsure	11	7.24
Total	152	100
14. What is your perception of AI's contribution to personalized medicine through bioinformatics? Select one)	Frequency	%
- Very High Contribution	46	30.26
- Moderate Contribution	55	36.18
- Low Contribution	8	5.26
- No Contribution	23	15.13
- Unsure	20	13.16
Total	152	100
15. What do you believe could be the next breakthrough in AI for bioinformatics and computational biology? (Select one)	Frequency	%
- Drug Discovery	19	12.50
- Disease Prediction and Diagnosis	55	36.18
- Personalized Medicine	33	21.71
- Genomic Sequencing and Analysis	23	15.13
- Other	22	14.47
Total	152	100

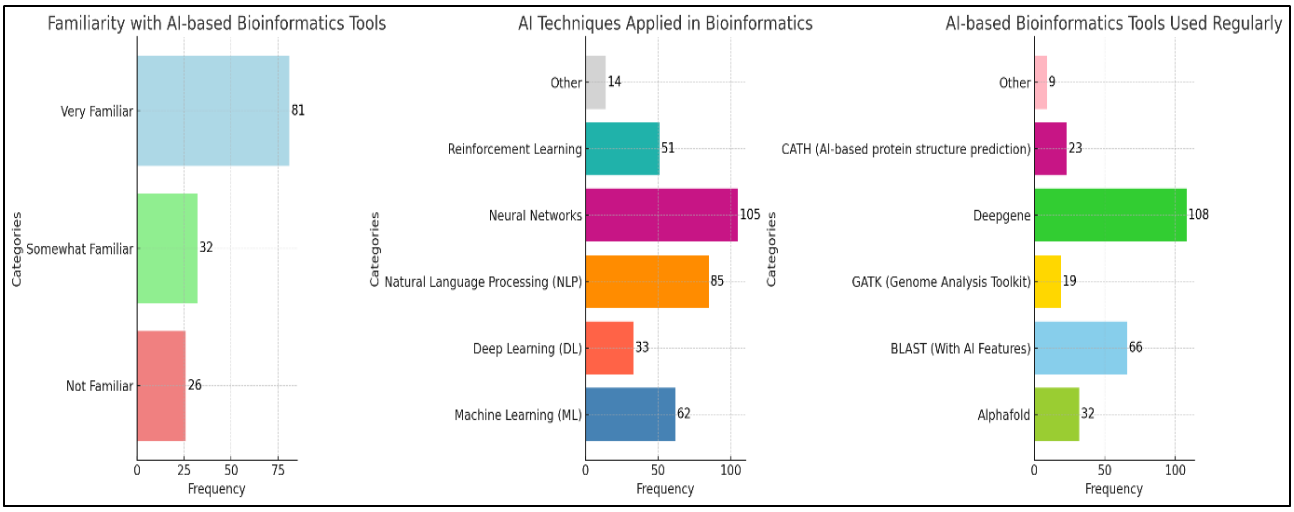


Fig.7. Future Trends and Opportunities

4.8 Thematic Analysis: Role of Artificial Intelligence in Information Extraction Systems

AI plays a pivotal role in transforming IE systems by enabling automation, improving accuracy, and streamlining processes. This thematic analysis investigates the impact of AI on IE systems, drawing insights from 12 respondents. Each respondent provided valuable feedback on the benefits, challenges, and future potential of AI in the field of information extraction. The results generated from this feedback are subsequently discussed to provide a comprehensive understanding of the current and future landscape of AI in IE systems.

4.8.1 Major Themes (MT) and Sub-Themes (STs)

Through conducting detailed interviews with a total of 12 respondents (designated as R-1 to R-12), the study has successfully identified several key themes and sub-themes. The classification of these findings is presented below:

4.8.2 MT1. The Impact of AI on Efficiency and Accuracy in Information Extraction

For this major theme (MT1), the following sub-themes were generated:

- Automation of Extraction Processes: ST1
- Improved Accuracy and Precision: ST2

Regarding the respondents' feedback, the actual outputs from R1, R2, and R5, covering the given main theme and sub-themes, are as follows:

- Respondent 1: "AI has been a game-changer in automating our document processing. Tasks that would take hours are now done in minutes."
- Respondent 2: "The AI tools we have implemented have significantly reduced the need for manual work. It is also allowing us to focus on more value-added activities."
- Respondent 5: "AI does not just automate tasks, but it also improves the accuracy of extractions. We now get more reliable data without human errors."

The theme of AI enhancing efficiency and accuracy was prominently reflected in the responses, as 10 out of the 12 respondents highlighted the automation and improvements in accuracy that AI has brought to their workflows.

4.8.3 MT2. AI Techniques Applied in Information Extraction

For this second major them or MT2, the following sub themes were generated.

- Machine Learning [22] Algorithms
- Natural Language Processing (NLP)
- Neural Networks

Regarding the respondents’ feedback, here are actual output from the R3, R4, and R6 covering the given main theme and sub themes.

- Respondent 3: "We are deeply reliant on machine learning models for categorizing data. The ability to train the system with new examples allows us to keep improving the extraction process with true letter and spirit."
- Respondent 4: " As per my view, Natural Language Processing is key in our information extraction, especially for unstructured text data. We can now extract meaningful insights from a vast amount of text data which would be beneficial in different tasks."
- Respondent 6: "Deep learning, especially neural networks has facilitated us to tackle more complex tasks like sentiment analysis within the data."

It is noteworthy that, for the second major theme and its associated sub-themes, nine respondents utilised Machine Learning algorithms, eight referred to the application of NLP for text-based extraction, and seven emphasised the significance of Neural Networks.

4.8.4 MT3: Challenges of Implementing AI in Information Extraction Systems

The third major theme identified through the interviews with respondents is titled "Challenges of Implementing AI in Information Extraction Systems." The associated sub-themes are as follows:

- Data Quality Issues
- Resource Intensity
- Model Complexity

Regarding the respondents' feedback, here is the actual output from the R7, R9, and R10 covering the given main theme and sub themes.

- Respondent 7: "The biggest challenge we face is data quality. If the data is noisy or incomplete, AI models struggle to extract accurate information."
- Respondent 9: "We require a lot of computing power to run our AI models. Therefore, it is among the significant cost factors that not everyone can afford."
- Respondent 10: "AI models are complex and require constant tweaking. Understanding what goes wrong when they fail can be quite challenging."

In terms of total feedback, it was observed that 7 respondents faced data quality issues, 8 respondents highlighted the resource intensity required for AI implementation, and 5 respondents mentioned the model complexity as a challenge.

4.8.5 MT4: Future Trends and Opportunities of AI in Information Extraction

This is the fourth major subject, covering the following sub-themes.

- Big Data Integration
- Real-time Information Extraction
- Improved Accuracy and Personalization

Regarding the respondents' feedback, here are actual output from the R8, R11, and R12 covering the given main theme and sub themes.

- Respondent 8: "I believe AI will increasingly integrate with big data systems, which will allow for better analysis and extraction across large datasets."
- Respondent 11: "Real-time information extraction is the future. With AI, it will be possible to process and extract data from live feeds."
- Respondent 12: "I see AI improving over time, leading to more personalized extraction systems tailored to specific industries or needs."

For this theme and associated sub-themes, overall, eight respondents mentioned the potential for big data integration in the future, seven highlighted real-time information extraction as a significant opportunity, and nine believed that AI will lead to improved accuracy and personalization, respectively. The overall distribution of the respondents' feedback using the sub-themes is presented in Figure 8.

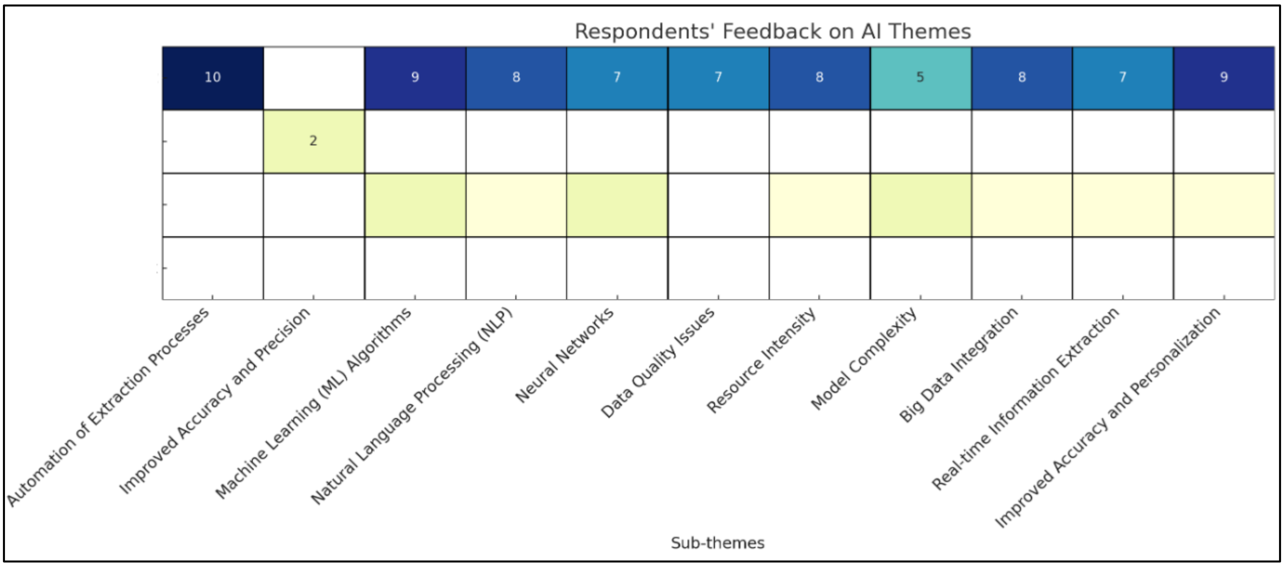


Fig.8. Sub-Themes as Per Respondents' View

5. Conclusion and Policies

This study used frequency distribution and thematic analysis to examine AI's role in bioinformatics and information extraction systems. The survey of 152 respondents revealed widespread familiarity with AI tools, particularly reinforcement learning, neural networks, and NLP. Key AI tools like CATH and DeepGene are commonly used in bioinformatics, with AI aiding in gene expression studies, biomarker identification, and protein structure prediction. However, challenges remain, including issues with AI model explainability, unstructured data, standardisation, and limited training data. Thematic analysis from 12 respondents highlighted AI's impact on improving efficiency and accuracy in information extraction, especially in data categorisation and text analysis. Key challenges include data quality, high resource demands, and model complexity. Looking ahead, AI is expected to integrate with big data, enable real-time extraction, and improve accuracy and personalisation. Policy recommendations include promoting data standardisation, ensuring data privacy, and fostering ethical AI practices to reduce bias. Investment in computing infrastructure and affordable cloud-based solutions is crucial for supporting AI applications. Additionally, AI models should be transparent and explainable, facilitating informed decision-making in research and healthcare. More funding should focus on AI-driven tools for real-time big data analysis.

References

[1] Lewis, J., & Bartlett, A. (2013). Inscribing a discipline: Tensions in the field of bioinformatics. *New Genetics and Society*, *32*(3), 243-263. https://doi.org/10.1080/14636778.2013.773172

[2] Zakarya, M., Rahman, I. U., Dilawar, N., & Sadaf, R. (2011). An integrative study on bioinformatics computing concepts, issues and problems. *International Journal of Computer Science Issues (IJCSI)*, *8*(6), 330. https://www.proquest.com/openview/13eb7a7dafc9c65ffa43bc9f09e96e7c/1?cbl=55228&pq-origsite=gscholar

[3] Kashyap, H., Ahmed, H. A., Hoque, N., Roy, S., & Bhattacharyya, D. K. (2015). Big data analytics in bioinformatics: A machine learning perspective. *arXiv preprint arXiv:1506.05101*. https://doi.org/10.48550/arXiv.1506.05101

[4] Can, T. (2014). Introduction to bioinformatics. *miRNomics: MicroRNA biology and computational analysis*, 51-71. https://doi.org/10.1007/978-1-62703-748-8_4

[5] Gauthier, J., Vincent, A. T., Charette, S. J., & Derome, N. (2019). A brief history of bioinformatics. *Briefings in bioinformatics*, *20*(6), 1981-1996. https://doi.org/10.1093/bib/bby063

[6] Maojo, V., & Kulikowski, C. A. (2003). Bioinformatics and medical informatics: collaborations on the road to genomic medicine? *Journal of the American Medical Informatics Association*, *10*(6), 515-522. https://doi.org/10.1197/jamia.M1305

[7] Lee, D., & Yoon, S. N. (2021). Application of artificial intelligence-based technologies in the healthcare industry: Opportunities and challenges. *International journal of environmental research and public health*, *18*(1), 271. https://doi.org/10.3390/ijerph18010271

[8] Abbas Khan, M., Khan, H., Omer, M. F., Ullah, I., & Yasir, M. (2024). Impact of artificial intelligence on the global economy and technology advancements. In *Artificial General Intelligence (AGI) Security: Smart Applications and Sustainable Technologies* (pp. 147-180). Springer. https://doi.org/10.1007/978-981-97-3222-7_7

[9] Hossain, K. A. (2023). Analysis of present and future use ofartificial intelligence (ai) in line of fouth industrial revolution (4ir). *Scientific Research Journal*, *11*, 1-50. http://dx.doi.org/10.31364/SCIRJ/v11.i8.2023.P0823954

[10] Oosthuizen, R. M. (2022). The fourth industrial revolution–smart technology, artificial intelligence, robotics and algorithms: Industrial psychologists in future workplaces. *Frontiers in artificial intelligence*, *5*, 913168. https://doi.org/10.3389/frai.2022.913168

[11] Keedwell, E., & Narayanan, A. (2005). *Intelligent bioinformatics : the application of artificial intelligence techniques to bioinformatics problems*. Wiley. https://cir.nii.ac.jp/crid/1130000794616810624

[12] Jirotka, M., Lee, C. P., & Olson, G. M. (2013). Supporting scientific collaboration: Methods, tools and concepts. *Computer Supported Cooperative Work (CSCW)*, *22*, 667-715. https://doi.org/10.1007/s10606-012-9184-0

[13] Outeiral, C., Strahm, M., Shi, J., Morris, G. M., Benjamin, S. C., & Deane, C. M. (2021). The prospects of quantum computing in computational molecular biology. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, *11*(1), e1481. https://doi.org/10.1002/wcms.1481

[14] Ezziane, Z. (2006). Applications of artificial intelligence in bioinformatics: A review. *Expert Systems with Applications*, *30*(1), 2-10. https://doi.org/10.1016/j.eswa.2005.09.042

[15] Nicolas, J. (2020). Artificial intelligence and bioinformatics. *A Guided Tour of Artificial Intelligence Research: Volume III: Interfaces and Applications of Artificial Intelligence*, 209-264. https://doi.org/10.1007/978-3-030-06170-8_7

[16] Jamialahmadi, H., Khalili-Tanha, G., Nazari, E., & Rezaei-Tavirani, M. (2024). Artificial intelligence and bioinformatics: a journey from traditional techniques to smart approaches. *Gastroenterology and Hepatology from Bed to Bench*, *17*(3), 241. https://doi.org/10.22037/ghfbb.v17i3.2977

[17] Valentini, G., Tagliaferri, R., & Masulli, F. (2009). Computational intelligence and machine learning in bioinformatics. In (Vol. 45, pp. 91-96): Elsevier. https://doi.org/10.1016/j.artmed.2008.08.014

[18] Shree, D. V., Kumar, B. S., Ganesan, V., Umadevi, N., Sivaraman, M., & Sumitha, J. (2024). The Role of Artificial Intelligence and Machine Learning Methodologies in Bioinformatics. 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 194-201. https://doi.org/10.1109/ICAAIC60222.2024.10575828

[19] Karunarathna, I., Gunasena, P., Hapuarachchi, T., & Gunathilake, S. (2024). Comprehensive data collection: Methods, challenges, and the importance of accuracy. *Uva Clinical Research*, 1-24. http://dx.doi.org/10.13140/RG.2.2.13134.47689

[20] Onwuegbuzie, A. J., & Leech, N. L. (2005). On becoming a pragmatic researcher: The importance of combining quantitative and qualitative research methodologies. *International journal of social research methodology*, *8*(5), 375-387. https://doi.org/10.1080/13645570500402447

[21] Taherdoost, H. (2022). What are different research approaches? Comprehensive review of qualitative, quantitative, and mixed method research, their applications, types, and limitations. *Journal of Management Science & Engineering Research*, *5*(1), 53-63. http://doi.org/10.30564/jmser.v5i1.4538ff.ffhal-03741840

[22] Ahmed, U., Zin, M., & Majid, A. (2016). Impact of Intention and Technology Awareness on Transport Industry's E-service: Evidence from an Emerging Economy. 산경연구논집 (IJIDB), 7 (3), 13-18. In. https://db.koreascholar.com/article/Detail/316838

[23] Smith, K. A., Morassaei, S., Ruco, A., Bola, R., Currie, K. L., Cooper, N., & Di Prospero, L. (2022). An evaluation of the impact for healthcare professionals after a leadership innovation fellowship program. *Journal of Medical Imaging and Radiation Sciences*, *53*(4), S137-S144.http://doi.org/ 10.1016/j.jmir.2022.09.004