

UNLOCKING SECRETS: BIOINFORMATICS' IMPACT ON FORENSIC BIO-EXAMINATIONS

Joe H Smith¹ and M Singh²

¹University of South Africa

²University of the Free State

ABSTRACT

Bioinformatics (BI) is a diverse and interdisciplinary domain that studies biological data through computational techniques. Its rapid technological progress has led to a significant increase in biological data production, sparking a notable expansion in BI research and application. In recent years, the convergence between BI and forensic science has become increasingly evident, with the growing use of BI tools and techniques in forensic investigations. BI is not only extensively used in biomedical research to analyse biological data and forecast gene regulation but also in the development of drugs targeting disorders linked to epigenetic alterations. It merges principles from various fields like biology, computer science, and statistics to process extensive omics data, create databases, predict structures, and model metabolic processes. The impact of BI on biomedical research is substantial, aiding in data processing, prediction, and visualisation. Moreover, BI is increasingly important in analysing data generated through Sanger-based electrophoresis methods and massively parallel sequencing in forensic DNA analysis, demonstrating its practical application in the field. Therefore, BI plays a critical role in forensic identification, searching databases side by side, forensic DNA phenotyping and biographic ancestry, forensic genetic genealogy, and forensic epigenomics. An overview of different bioinformatic methodologies employed in forensic applications, exploring approaches for managing data and investigating trend analysis strategies within forensic contexts, is discussed. Additionally, it investigates data management strategies. The primary objective was to explore how BI facilitates forensic BI's use, integration, and impact. Utilising a systematic literature review and content analysis, the study examined peer-reviewed articles, conference proceedings, and reports from critical databases like Google Scholar and Scopus, focusing on publications from 2010 to 2024. Key findings highlight the significant advancements in genomic sequencing technologies and their application in personalised therapeutics. Forensic scientists can improve forensic exams' accuracy, reliability, and efficiency by using BI tools, metrics, and efficient data management procedures.

KEYWORDS

Deoxyribonucleic acid (DNA) data, bio-examinations, bioinformatics, forensic examinations, massive parallel sequencing,

1. INTRODUCTION

Forensic science is essential in contemporary criminal investigations, as it offers invaluable evidence and enhances the understanding of legal proceedings. Advances in high-throughput sequencing methodologies and extensive biomedical investigations have resulted in an unparalleled accumulation of biological data, facilitating computer analysis based on data. Computational approaches can process data, forecast, and examine prospective patterns within extensive biological datasets, unlike conventional wet laboratory-based research, which often requires significant labour and time investment. The number of candidates for further experimental validation can be reduced, thereby optimising the research or examination process [1].

BI is a collaborative and interdisciplinary domain that harnesses computational techniques to study biological data. The rapid advancement of technology has led to a significant increase in biological data production, sparking a remarkable expansion in BI research and application. The convergence of BI and forensic research has opened up new avenues for examining complex biological data, detecting genetic markers, and predicting forensic inquiries. BI is the science of how information is generated, transmitted, received, stored, processed, and interpreted in biological sciences or the application of information science to Biology [2-10]. BI has evolved significantly, merging multiple academic disciplines, including mathematics, statistics, computer science, engineering, and biotechnology. Moreover, it has significantly impacted various biofields such as deoxyribonucleic acid (DNA) genomics (including forensic examinations, metagenomics, proteomics, and epigenomics), microbiomics, and exposomics [1-10]. This collaborative approach underscores the shared purpose and unity between bioinformatics and forensic science, paving the way for innovative solutions in both fields.

The implementation of highly massive parallel sequencing (sometimes called next-generation sequencing) methods has significantly increased the number of DNA sequences stored in publicly accessible genetic (direct-to-consumer) databases directly accessible to consumers [11]. This surge in data generation necessitates significant computational resources, including storage and organisation, to enable convenient access and utilisation. Computers and specialised software have become indispensable tools in biologists' repertoires. The use of computers and BI is crucial in nearly all modern biological research projects and forensic examinations, whether they involve standard DNA or protein sequence analysis or deriving significant insights from large biological datasets. At its core, BI integrates molecular biology and computer informatics to effectively handle, examine, and archive genetic data. This practical application of bioinformatics in forensic science offers a promising future, where the use of BI tools and techniques can revolutionise the accuracy, efficiency, and reliability of forensic examinations.

DNA profiling is used to identify an individual, and together with bioinformatics, DNA profiling advantages can be displayed. BI enables one to store and analyse individuals' genetic information; an example is the DNA Database. DNA databases are specialised software applications that have emerged from BI [12]. This application has garnered significant significance among biologists and is increasingly recognised in forensic research. DNA databases are utilised in a wide range of applications, encompassing the analysis of genetic illnesses, the implementation of genetic fingerprinting techniques in forensic investigations, and the practice of tracing genealogy. It has become evident that combining DNA profiling techniques and database matching is a powerful tool in forensic casework. The DNA profile results produced from biological samples are entered into forensic databases during the matching process to evaluate the probability of a match and ascertain its relevance [3].

To comprehend the complexities of BI tools, life scientists not actively involved in BI programmes must comprehend fundamental principles related to BI science. This understanding is of utmost importance to mitigate the potential for misuse and erroneous interpretations of findings [13-16].

This paper examines BI techniques in forensic science, focusing on enhancing the efficiency of forensic examinations. It also explores approaches for managing data and investigates trend analysis strategies within forensic contexts.

1.1. Aim and Objectives of the Study

This study aims to comprehensively analyse and understand BIs' role in advancing and implementing forensic BI. The study explores how bioinformatics is revolutionising the field by integrating forensic DNA and genomics.

The objectives are:

- To analyse the use, integration, and impact of forensic BI.
- To assess the challenges, ethical and privacy concerns with forensic BI

2. METHODOLOGY

The methodology employed a systematic literature review focusing on peer-reviewed sources in forensic BI. Data sources included databases like Google Scholar, PubMed, and Scopus. The search strategy used keywords related to forensic bioinformatics, ethics and privacy, forensic bioinformatics, accreditation DNA, and challenges forensic biometrics, limited to English articles from mainly 2010 to 2024. Inclusion criteria covered the more recent and peer-reviewed articles on bioinformatics in forensic applications and related accreditation ethical and privacy concerns. Exclusions included non-peer-reviewed articles, opinion pieces, and non-English papers. The selection process involved screening titles/abstracts and full-text reviews meeting inclusion criteria and providing helpful information. Data analysis used content analysis to identify themes and patterns relevant to technological advancements, impact, challenges, and ethical and privacy considerations with forensic biometrics.

3. LITERATURE REVIEW

3.1. DNA Analysis and Sequencing

Forensic DNA analysis is a fundamental component of criminal investigations. It provides insights into individual identification, kinship analysis, and genetic markers relevant to criminal investigations. Various BI tools (e.g. Assembler, MacVector, BLAST, Bowtie, BWA, and USAT) are widely employed for tasks such as sequence comparison and alignment, variation calling, allele size conversion, plotting, and visualisation of genomic data and comparing DNA profiles with databases. Assembler is a plugin for MacVector that provides comprehensive sequence assembly functionality. Assembler is fully integrated into MacVector and allows the managing of sequencing data using the familiar MacVector style. Primer design can be configured [17-20].

BI methods can also be used for population genetics models, maximum likelihood estimation, and evolutionary relationships. Based on genetic markers, these methods determine ancestral origins and population affiliations. These analysis techniques facilitate identifying individuals and enhance comprehension of population migrations and genetic variation within forensic contexts. Moreover, BI is employed in multiple phases of DNA sample processing, encompassing isolation, quantification, fragment separation analysis, and allele identification and interpretation.

The methods and instrumentation used to process forensic samples have advanced far beyond what traditional binary mixture interpretation methods could do to identify the contributors to forensic DNA profiles. Conventional DNA analysis approaches face difficulties when dealing with intricate combinations, such as those derived from "touch DNA" items or post-coital sexual assault samples that involve several contributors and stochastic data. When it comes to forensic

DNA mixture analysis, binary procedures need assistance precisely evaluating low-level mixtures with dropout because they need exact mathematical validation and can only get close to the weight of evidence.

Through probabilistic genotyping methods, BI application includes uncertainty in likelihood ratio calculations when analysing intricate mixtures. Probabilistic genotyping software, like BP Sentry, DNASTatistX, EuroForMix, LRmix, Mixolydian, STRmix, and TrueAllele, is essential for criminal investigations because it helps link DNA samples found at crime scenes (mainly where several persons have contributed to the DNA found on the sample) with a specific person of interest (POI). This BI software facilitates genetic analysis in intricate situations where samples exhibit partial degradation or encompass DNA from numerous individuals.

Most deconvolution solutions have probabilistic characteristics as opposed to deterministic ones. The software used for probabilistic genotyping computes a likelihood ratio (LR) to quantify the level of evidence, evaluating the evidence about various pairs of hypotheses. For example, it calculates the likelihood that the DNA is associated with the suspect compared to the likelihood that it is associated with an unidentified, unrelated person. Furthermore, using probabilistic genotyping software requires the presence of a suspect's DNA to facilitate the comparison of results. In contrast, Mixoldian is deterministic. The software used for probabilistic genotyping computes a likelihood ratio (LR) to quantify the level of evidence, evaluating the evidence about various pairs of hypotheses. For example, it calculates the likelihood that the DNA is associated with the suspect compared to the likelihood that it is associated with an unidentified, unrelated person. Furthermore, using probabilistic genotyping software requires the presence of a suspect's DNA to facilitate the comparison of results. Probabilistic genotyping software uses iterative Bayesian inference networks. Mixolydian's technique, on the other hand, introduces a novel mathematical formula that can perform a comprehensive deconvolution procedure to produce the genotypes of individual contributors directly. Probabilistic genotyping software necessitates the presence of a suspect's DNA to compare results. In contrast, Mixolydian provides the precise genotype, which may be swiftly inputted into CODIS and other databases to facilitate potential match searches.

The Human Genome Project has facilitated the development of novel technologies that aid in forensic examinations and kinship analysis by identifying evidence sources at crime scenes and unidentified human remains. Massively parallel sequencing (MPS) or next-generation sequencing (NGS) is an innovative technology that provides automated, high-throughput capabilities. It enables the simultaneous analysis of millions of genetic markers per sample while maintaining sample consumption levels equivalent to those of existing forensic methods. NGS, in conjunction with its supplementary techniques, provides numerous benefits. This technology facilitates the examination of a considerably greater quantity of genetic markers per sample, spanning from tens of thousands to millions. Incorporating NGS and its related approaches enhances the thoroughness and precision of kinship analysis. It also expands the scope of forensic genetics to encompass a broader array of intricate situations, thus propelling the discipline of forensic science forward [21].

Furthermore, the progress made in NGS technologies has facilitated the high-throughput sequencing of DNA samples, enhancing the ability to analyse intricate DNA mixtures and degraded samples frequently seen in forensic casework. Instead of using older methods, NGS are a better and more affordable way to find tandem repeat alleles as sequence variations or haplotypes [11]. These BI tools process sequencing datasets to find tandem repeat haplotypes [22-24]. The software program can detect alleles within tandem repeats, encompassing length- and sequence-based variations. As denoted by its sequence, the short tandem repeat (STR) haplotype offers comprehensive data encompassing the number of repetitions of a fundamental

pattern and further variations, like single nucleotide polymorphisms and insertions/deletions, if they are present. The utilisation of this integrated strategy serves to augment the precision and comprehensiveness of the data acquired through tandem repeat analysis, employing NGS technology and bioinformatic tools [20].

Moreover, bioinformatic tools like the Universal STR Allele Toolkit (USAT) enable the thorough examination and interpretation of tandem repeat haplotypes. USAT has various functionalities, including, but not limited to, converting length-based and sequence-based alleles, nucleotide comparison of tandem repeat haplotypes, an atlas of allele distributions, data formatting, interactive data filtering, and visualisation in parallel computing [20].

3.2. Examination of Phylogenetics and Ancestry

Phylogenetic analysis and ancestry inference are significant in forensic genetics, particularly in human remains or an unidentifiable population [25]. Bioinformatic methods, such as Bayesian inference, maximum likelihood estimation, and population genetics models, reconstruct and infer evolutionary relationships and ancestral origins and determine population affiliations based on genetic markers. These analysis techniques facilitate identifying individuals and enhance comprehension of population migrations and genetic variation within forensic contexts [25].

BI techniques can facilitate phylogenetic reconstructions of polyploids, offering enhanced functionality and usability. The application of BI can facilitate the allocation of alleles from polyploids to diploid sub-genomes in multi-locus sequence datasets. In addition, the BI can provide a consensus methodology for evaluating the reliability of phylogenetic reconstructions. BI highlights the efficacy and precision that can be accomplished through this methodology in conducting phylogenetic investigations about polyploid genomes [26].

While anthropological and historical substrates may need more genetic information due to incomplete profiles, this constraint does not make them ineffective. Using BI tools like genotyping extraction techniques, researchers can make necessary inferences about people's identification, biogeographic heritage, maternal and paternal lineages, and observable physical traits [25].

Using BI and incorporating biogeographical ancestry inference into conventional forensic DNA profiling techniques is a significant advancement. It is crucial in human forensic inquiries, mainly when no database match or eyewitness testimony exists. BI with biogeographical ancestry inference uses population-specific genetic variants to ascertain an individual's biogeographical origin, assisting authorities in identifying unidentified persons. This forensic BI-driven testing utilises progress in comprehending human genomic variation and molecular biology, making a substantial contribution to the advancing capabilities of forensic science [27].

3.3. The Application of Proteomics and Metabolomics in Forensic Criminology

The progress made in proteomics drive examination using BI has created body fluid identification techniques essential for forensic purposes. We employ BI proteomics technologies to detect and identify proteins in diverse bodily fluids frequently seen at crime scenes or on body surfaces. Preserving DNA integrity for subsequent analysis is a notable advantage of BI proteomics' non-destructive nature [28,29].

Proteomic and metabolomic techniques using BI significantly contribute to forensic investigations in toxicology, drug analysis, postmortem research, and DNA analysis. In conjunction with bioinformatic pipelines, mass spectrometry facilitates the detection and

characterisation of proteins, peptides, metabolites, and medicinal molecules within biological specimens. Analysing these types of data is essential for elucidating deaths associated with drug consumption, identifying indicators of physiological changes, and linking chemical profiles to forensic scenarios. [30].

3.4. The Field of Microbial Forensics and BI

The human microbiome, which consists of a varied collection of microbial cells primarily made up of bacteria, exists in numerous areas, including the gastrointestinal tract, skin and oral cavity. This microbial community holds significant promise in forensic research, serving crucial functions in postmortem examination, geographical characterisation, sexual assault inquiries, and personal identification [31]. Personal variables, including lifestyle, genetics, health status, and geographical origin, significantly influence microbiome samples' distinctive traits and forensic value. Microbiome analysis using a forensic tool is possible due to advances made in NGS and BI [32]. Microbiome samples are employed for several purposes, including human identification, measuring postmortem intervals, and determining geolocation. By utilising advanced technologies such as MPS and long-read sequencing, in conjunction with BI tools and machine learning models, it becomes possible to analyse and describe individual samples of the human microbiome efficiently and accurately [33]. Metagenomics Analysis Software (MEGAN) allows visual and quantitative exploration of microbiomes on a desktop computer. The information is annotated concurrently to cover taxonomic and functional composition across various databases [34].

The discipline of microbial forensics is an expanding study area combining microbiology, BI, and forensic science. Microbial genomes, microbial communities, and the origins of biological agents in situations of bioterrorism, infectious disease outbreaks, or environmental pollution episodes are analysed using bioinformatic tools and databases. Microbial forensics studies rely heavily on metagenomic NGS, genome assembly, and phylogenetic profiling techniques with BI [35].

3.5. The Significance of Metrics in Forensic Research and Examinations

Metrics function as numerical indicators used to appraise forensic procedures, evaluate analytical effectiveness, and guarantee the dependability of forensic findings. The practical interpretation of DNA evidence and determining the relevance of genetic matches in forensic DNA analysis relies on metrics such as match probabilities (e.g., Combined Probability of Inclusion, random match probabilities, and allele dropout rates). Similarly, the evaluation of bioinformatic analysis performance involves the assessment of various metrics such as alignment accuracy, variant calling sensitivity, specificity, and false discovery rates. These metrics indicate the reliability and precision of computational techniques utilised in forensic investigations.

Metrics such as likelihood ratios (LR) are used in forensic casework to evaluate the probative value of DNA evidence. These metrics quantify the probabilities of observing a specific DNA profile under various hypotheses, such as inclusion versus exclusion. Integrating population databases and allele frequency data into LR calculations augments the evidential potency of DNA profiles, hence assisting forensic experts and legal professionals in their decision-making endeavours [10, 33]. Utilising suitable reference databases for the computation of match probability is imperative, given the influence of genetics on the interpretation of DNA evidence [36-38].

Forensic laboratories can assess the performance of bioinformatic tools by comparing them to reference datasets, standards, and proficiency testing samples [37-49]. BI helps evaluate sensitivity, specificity, accuracy, precision, and computing efficiency. The reliability and

reproducibility of bioinformatic studies in forensic contexts are ensured through the implementation of quality control methods, validation protocols, and adherence to international standards such as ISO/IEC 17025 [37-49].

Implementing effective data management strategies is of utmost importance in forensic BI to effectively manage the substantial quantities of genomic, proteomic, metabolomic, and metadata produced during forensic investigations. Effective data management strategies in forensic research necessitate incorporating crucial elements such as secure data storage, standardised data formats, metadata annotation, and data sharing protocols.

Forensic laboratories should establish and enforce secure data storage systems, backup protocols, access controls, and encryption measures to protect sensitive forensic data from unauthorised access, breaches, or tampering [37]. Adhering to data protection rules and forensic data preservation protocols is imperative to maintaining forensic evidence's integrity, confidentiality, and legal admissibility.

3.6. Analysis of Trends and Identification of Patterns

Temporal trends, anomalies, and significant insights can be extracted from forensic information through trend analysis and pattern recognition tools, empowering forensic scientists. In many forensic settings, trend analysis, forensic data mining, and pattern identification are facilitated through data visualisation tools, machine learning algorithms, statistical modelling approaches, and network analysis methodologies [37, 50].

Machine learning algorithms, such as, clustering algorithms, random forests, and deep learning models, are commonly used in forensic datasets for tasks like pattern recognition, evidence type classification, species identification, and prediction of forensic outcomes. Machine learning models in forensic applications can improve robustness and generalisation by utilising feature selection methods, cross-validation tactics, and model interpretability tools [37].

In forensic datasets, various network analysis techniques, including social network analysis, network analysis, and semantic network analysis, are employed to uncover latent connections, links, and interactions. Forensic experts can utilise visual analytics platforms, graph visualisation tools, and interactive dashboards to examine intricate data connections, detect areas of high interest in forensic investigations, and effectively convey analytical results to stakeholders, investigators, and legal professionals.

3.7. Challenges Facing Forensic Bioinformatics

biological Forensic BI encounters various technological and methodological obstacles that affect its operational efficiency and overall efficacy. One vital obstacle is effectively handling the vast data produced during forensic investigations. Due to the progress made in sequencing technologies and data production methodologies, forensic laboratories frequently encounter extensive datasets that necessitate storage, organisation, and rapid retrieval for analysis. One further significant obstacle is the velocity of data analysis within forensic bioinformatics. Conventional approaches often need help to efficiently handle extensive datasets, leading to potential delays in the delivery of outcomes. It is crucial to remember that efficient analysis is not just a matter of operational efficiency but a critical factor in forensic investigations. It facilitates prompt replies and informed decision-making, which can impact the outcomes of forensic investigations [37, 51-53].

Computational costs further hinder forensic bioinformatics. The processing, analysis, and interpretation of intricate biological data can impose significant computational requirements, straining resources such as hardware infrastructure and software tools. Forensic laboratories and researchers are constantly concerned with effectively managing these expenses while ensuring precision and dependability [37, 51-53].

These issues necessitate the implementation of novel strategies in the domains of data management, analysis methods, and computational infrastructure. Introducing scalable storage solutions, parallel computing techniques, and optimised algorithms promises to enhance data handling and analysis speed. Moreover, the progress in communication systems (e.g. Voice over Internet Protocol – VoIP) is often deployed to facilitate effective communication. Developing customised solutions for these difficulties necessitates the collaboration of bioinformaticians, computer scientists, forensic professionals, and policymakers. Forensic bioinformatics will be advanced through research and development endeavours that prioritise data processing, enhance analysis tools, and implement cost-effective computing solutions. These efforts will improve forensic investigations and outcomes [37, 51-53].

Consequently, forensic managers or decision-makers must choose the best BI solution and information technology infrastructure from a vast list of laboratory and testing method options. They often need to gain the right skills and rely on help from outside sources, which could be biased towards certain suppliers and might only sometimes be the best fit for their needs. Procuring a BI system and information technology infrastructure with sustained maintenance to keep them operating can be expensive [37, 51-53].

Cloud computing and distributed computing architectures present exciting prospects for diminishing computational expenses and augmenting scalability inside forensic bioinformatics procedures, paving the way for a more efficient and effective future. Developing customised solutions for these difficulties necessitates the collaboration of bioinformaticians, computer scientists, forensic professionals, and policymakers. Forensic bioinformatics will be advanced through research and development endeavours that prioritise data processing, enhance analysis tools, and implement cost-effective computing solutions. These efforts will improve forensic investigations and outcomes [37, 51-53].

Consequently, forensic managers or decision-makers must choose the best BI solution and information technology infrastructure from a vast list of laboratory and testing method options. They often need to gain the right skills and rely on help from outside sources, which could be biased towards certain suppliers and might only sometimes be the best fit for their needs. Procuring a BI system and information technology infrastructure with sustained maintenance to keep them operating can be expensive [37, 51-53].

3.8. Ethical and Privacy Issues

Genetic and biological information in forensic bioinformatics is private and sensitive, and ethics and privacy problems must be managed appropriately. Legal, scientific, and moral steps must be taken together in a multifaceted way to protect people's privacy while using DNA databases and bioinformatic analyses to help with forensic investigations. Rules or frameworks to protect people's privacy must be developed when DNA databases and bioinformatic studies are used in forensic investigations. These rules should cover many topics, like collecting, storing, sharing, and controlling who can see and analyse data [37, 55-64].

One important aspect that demands attention is enforcing stringent ethical standards and legal rules in data handling and privacy security. These rules necessitate ensuring that individuals are

fully informed about the potential risks of collecting and retaining their DNA information. Robust encryption database storage and transmission security measures can effectively deter unauthorised access and data breaches. Furthermore, it is imperative to establish unambiguous rules for how forensic organisations, researchers, and law enforcement can access and share data. Access should be strictly limited to authorised personnel, and data sharing should only occur under stringent privacy and control regulations, thereby fostering transparency and accountability [37, 55-64].

Securing informed consent from individuals whose genetic data is being collected, stored, or analysed is not just a critical ethical issue in forensic bioinformatics; it is the cornerstone of ethical practice. Informed consent is not a mere formality but a fundamental right, especially involving biometric data and genetic records. Individuals must be fully aware of and agree to how their data will be used [37, 55-64].

Genetic and biological data must be kept safe from hacking or misuse. Forensic bioinformatics databases and laboratories must follow strict security rules and encryption methods to protect private data from hackers and others who should not have access to it [37, 55-64].

Clarity on the ownership and control of DNA data is not just a legal necessity but a moral imperative. For responsible forensic investigations, it is crucial to establish unambiguous rules and policies that clearly define who owns the data, how it can be shared, and how it can be used. This framework ensures fairness and transparency, which are vital elements in maintaining public trust [37, 55-64].

Genetic data can reveal intimate details about a person's health, traits, and ancestry, which can be misused. The intimate and private issues in forensic bioinformatics require acute attention and management. Removing personally identifiable information from data is standard practice to protect people's identities. However, improvements in data analytics and re-identification methods make it harder to stay anonymous, so strong de-identification methods are needed [37, 55-64].

When forensic DNA databases are created and kept up to date, they raise ethical questions about how much data should be collected, how long it should be kept, and who should be able to view it. Finding a balance between the need to investigate and people's rights to privacy and civil freedoms is complex and must be considered within the legal framework. BI algorithms and data interpretation methods must address biases and errors to ensure that forensic analyses are fair and accurate. Biases in databases, analysis tools, or ways of interpreting data can cause people to come to wrong conclusions or keep biases alive in society [37, 55-64].

It is essential to ensure that bioinformatic tools are used honestly and responsibly in criminal investigations. Thus, it means not using genetic profiling based on race or culture, only using data for legal reasons, and respecting human rights and dignity while the investigation is ongoing. To solve these ethical and privacy problems, bioinformaticians, forensic experts, policymakers, ethicists, and advocates must collaborate to create complete rules, guidelines, and ethical frameworks that encourage openness, responsibility, and respect for individual rights in forensic bioinformatics practices [37, 55-64].

4. RESULTS AND DISCUSSION

In an ideal world, new technologies would be tailored to meet the needs of forensics and bioresearch, ensuring that they can be used legally and effectively in forensic and bio laboratories. Effective technologies mean determining the worth of the evidence and adding a

criminalistic interpretation. Modern forensic DNA profiling is an excellent example of how scientific findings, technological growth, BI technology, and forensic use and interpretation can all come together.

At its core, BI seamlessly merges the realms of molecular biology with computer informatics, enabling the efficient management, analysis, and preservation of biological (genetic) data. This powerful tool, rooted in statistical and computational principles, is effective in various scientific fields and crucial in precisely identifying parent-child relationships and other familial connections, even in the challenging context of mass disasters.

Moreover, it is essential to highlight that BI is gaining significant traction in forensic statistics, particularly in assessing the statistical significance of DNA profiles [37]. Developing specialised software that is now essential for biologists and is receiving increasing recognition from forensic professionals marks the DNA database as a prime example of a BI application. This database is a vital tool in genetic analysis, forensic investigations, and kinship analysis, underscoring the adaptability and importance of BI in modern scientific and forensic spheres. DNA databases are an essential tool for forensic scientists, enabling them to solve crimes and aid in identifying unidentified bodies. BI facilitates comparing profiles within the DNA database, finding profiles that match each other, and identifying profiles with genetic similarities [37-39]. BI tools with a graphical user interface are being developed to facilitate DNA sequence comparisons, convert allele sizes, and visually represent genomic tandem repeat haplotypes [20,37].

Recently, BI tools have become adaptable and may be utilised in diverse domains, including forensics, epigenetics illness diagnostics, evolutionary genomics, and breeding research, to analyse tandem repeat haplotypes obtained from different sources effectively. The fundamental objective of BI is to enhance precision and effectiveness in examining and comprehending data. [37].

Efficient information communication technology is vital in modern scientific and technical fields, managing and analysing vast data for improved research and examination, outcomes, and decision-making. For example, VoIP, a broadband technology, revolutionises communication and collaboration, particularly in bioinformatics and forensics. It offers flexibility, cost-effectiveness, and advanced features compared to traditional cellular communication. VoIP enables seamless communication through various channels like calls, video conferences, and file sharing, facilitating remote collaboration, data sharing, and real-time discussions [41]. Its integration empowers forensic bioinformatics teams, streamlining operations, enhancing communication, and improving forensic processes' efficacy, thus benefiting scientific research and forensic practices.

It is essential to develop rules or frameworks to protect people's privacy when DNA databases and bioinformatic studies are used in forensic investigations. These rules should cover many topics, like collecting, storing, sharing, and controlling who can see and analyse data. The potential consequences of not enforcing stringent ethical standards and legal rules in data handling and privacy security are grave. It is imperative to ensure that individuals understand the privacy and other risks associated with collecting and retaining their DNA information. Robust encryption database storage and transmission security measures can effectively deter unauthorised access and data breaches [37. 55-64].

Implementing quality management systems, including developing policies and standing operating procedures that safeguard people's privacy and information safety, is essential when DNA databases and bioinformatic analyses are used in forensic investigations. Regular audits and evaluations of data handling can help identify and reduce privacy risks. To ensure that rules and guidelines are followed, it is also important to teach everyone involved how important it is to

protect privacy and consider ethics when conducting BI studies. Independent peer review (accreditation) ensures compliance with the quality management system that follows International Organisation for Standardisation (ISO) standards, and regulatory oversight should not be optional. This independent review and oversight ensure that all privacy, data integrity, data sharing, data use for its intended purpose, access, and security issues are adequately addressed when BI solutions are implemented [37-49].

5. CONCLUSION

The exponential progress in technology has precipitated a substantial surge in the generation of biological data, facilitating the broadening of studies and utilising BI. Extensive data from instruments and applications require efficient information communication technologies. There has recently been an observable convergence between BI and forensic science, as bioinformatic technologies are progressively employed in forensic investigations. Incorporating bioinformatic tools, metrics-driven procedures, streamlined data management practices, and sophisticated trend analysis techniques has fundamentally transformed forensic science. BI has empowered forensic laboratories and investigative agencies to address intricate obstacles to effectively processing evidence.

Biological investigation (BI) plays a vital role in biomedical research by facilitating the analysis of biological data, forecasting gene regulation, and discovering drugs for illnesses associated with epigenetic alterations. (BI) plays a substantial role in several analytical applications, including forensic identification, database searches, forensic DNA phenotyping, biographic ancestry determination, genetic genealogy, and epigenomic analysis. Integrating principles from various fields, such as biology, computer science, mathematics, physics, and statistics, makes extracting pertinent insights from biological data easier.

Furthermore, it is crucial to examine data derived from forensic DNA analysis methodologies, including Sanger-based electrophoresis and large-scale parallel sequencing. BI plays a substantial role in several analytical applications, including forensic identification, database searches, forensic DNA phenotyping, biographic ancestry determination, genetic genealogy, and epigenomic analysis. Integrating principles from various fields, such as biology, computer science, mathematics, physics, and statistics, makes extracting pertinent insights from biological data easier.

BI effectively manages substantial amounts of data derived from omics technologies. It carries out several functions, like the construction of databases, the analysis of genomes, the prediction of 3D structures, the discovery of drugs, clinical applications, and metabolic modelling. Moreover, business intelligence technologies play a vital role in epigenomic profiling and investigating the factors that drive epigenetic modifications. This article examines the intersection of business intelligence and forensic bioscience, delving into the utilisation of bioinformatic approaches in forensic contexts, strategies for managing data, analysis of trends, and the enhancement of forensic examination precision through the implementation of BI tools, metrics, and effective data management procedures.

Computational modelling may involve comprehensively examining live organisms and their environs, considering all molecular categories concurrently instead of separately analysing whole genomes, transcriptomes, or metabolomes. BI has dramatically augmented biologists' and forensic scientists' capacities in identifying individuals across diverse criminal and non-criminal scenarios. The DNA database is a vital BI tool, functioning as a comprehensive platform for managing, analysing, and storing genetic information. DNA databases are significant in forensic investigations as they match DNA profiles using statistical and computational techniques.

BI is paramount in investigating and using efficient data management methodologies. By utilising BI tools, metrics, and efficient data management techniques, forensic scientists can significantly improve forensic examinations' accuracy, reliability, and efficiency. These technological instruments facilitate the examination of intricate biological data, assisting in comprehending and verifying forensic evidence. By incorporating sophisticated metrics and effective data management rules, forensic analysis workflows can be strengthened, leading to enhanced investigative outcomes and improvements in court processes.

BI is essential for evaluating the effectiveness of forensic bioanalysis by comparing it to established parameters and production output. The extraction of necessary trend analysis data through BI enables the identification of early warning indications for deviations from established norms. The utilisation of this data not only facilitates the enhancement of decision-making processes but also empowers forensic scientists to optimise workflows and improve the overall outcomes of forensic examinations. In this context, the enhanced application of BI promotes a proactive stance in upholding quality standards and recognising opportunities for ongoing enhancement in forensic investigations.

When implementing forensic bioinformatics, careful consideration of ethical and privacy concerns is crucial. When implementing BI, solid laws, rules, and laboratory protocols must be carefully considered and supported. These include getting informed consent, ensuring data is private and safe, figuring out who owns and controls the data, and addressing the possibility of discrimination.

Forensic managers must consider the technological and methodological hurdles encountered in forensic bioinformatics. These challenges include effectively managing data volume, enhancing analysis speed, and addressing computational costs.

Forensic BI is on the verge of transforming the DNA profile-matching process in the future. Integrating informatics into forensic and genetic technology is of utmost importance for African countries, considering the worldwide scope of forensic DNA analysis. This integration will promote cooperation among regions and enable the sharing of crucial information required to strengthen criminal investigations, identify human remains, and aid funds in finding missing persons. Integrating informatics within this context will enhance forensic capabilities and improve outcomes in resolving intricate situations.

REFERENCES

- [1] F. Li et al., (2022), "Positive-unlabeled learning in Bioinformatics and Computational Biology: A brief review," *Brief. Bioinform.*, vol. 23, no. 1, p. bbab461, doi:10.1093/bib/bbab461.
- [2] R. Al Hakim et al., (2022), "Current evidence on BI role and digital forensics that contribute to forensic science: Upcoming threat," *J. Riset Rumpun Mat. Ilmu Pengetahuan Alam (JURRIMIPA)*, vol. 1, no. 1, pp. 25-32.
- [3] F. Li et al., (2022), "Positive-unlabeled learning in Bioinformatics and Computational Biology: A brief review," *Brief. Bioinform.*, vol. 23, no. 1, p. bbab461, doi:10.1093/bib/bbab461.
- [4] H. Li et al., (2020) "Modern deep learning in bioinformatics," *J. Mol. Cell Biol.*, vol. 12, no. 11, pp. 823-827, doi:10.1093/jmcb/mjaa030.
- [5] J. Gauthier et al., (2019), "A brief history of Bioinformatics," *Brief. Bioinform.*, vol. 20, no. 6, pp. 1981-1996, doi:10.1093/bib/bby063.
- [6] D. Mariano et al., (2020), "A brief history of bioinformatics told by data visualisation", In: Setubal, J.C., Silva, W.M. (eds) *Advances in Bioinformatics and Computational Biology*. BSB Lecture

Notes in Computer Science, vol 12558. Springer, Cham. https://doi.org/10.1007/978-3-030-65775-8_22

- [7] J. Ramsden, (2023), *Bioinformatics: An Introduction*. Springer Nature.
- [8] P. H. Russell et al., (2018), "A large-scale analysis of bioinformatics code on GitHub," *PLOS ONE*, vol. 13, no. 10, p. e0205898, doi:10.1371/journal.pone.0205898.
- [9] L. H. Santos and R. C. de Melo-Minardi, (2020). "A brief history of bioinformatics told by data visualisation" in *Advances in Bioinformatics and Computational Biology, Proceedings: 13th Brazilian Symposium on Bioinformatics, BSB 2020, São Paulo, Brazil, November 23-27*, vol. 12558. Springer Nature, 2020, Dec.
- [10] H. Susanto et al., (2020) "Bioinformatics: Progress, trends, and challenges," *Biochem. Biophys. Mol. Chem.*, pp. 231-250.
- [11] B. E. Slatko et al.,(2018), "Overview of next-generation sequencing technologies," *Curr. Protoc. Mol. Biol.*, vol. 122, no. 1, pp. e59-e59, doi:10.1002/cpmb.59.
- [12] A. N. S. T. Meilana and E. I. Auerkari, (2024), "DNA profiling, BI and databases in forensics: Human identification purposes," *J. Lifesci*, vol. 5, no. 1, pp. 49-63, doi:10.37899/journalllifesci.v5i1.1171.
- [13] R. A. C. dos Santos et al., (2022), "Original strategies for training and educational initiatives in bioinformatics. In frontiers in education," *Frontiers*, vol. 7.
- [14] D. Gatherer, (2020), "Reflections on integrating bioinformatics into the undergraduate curriculum: The Lancaster experience," *Biochem. Mol. Biol. Educ.*, vol. 48, no. 2, pp. 118-127, doi:10.1002/bmb.21320.
- [15] D. Mariano et al., (2019), "Introducing programming skills for life science students," *Biochem. Mol. Biol. Educ.*, vol. 47, no. 3, 288-295, doi:10.1002/bmb.21230.
- [16] A. T. Vincent, et al., (2018), "Implementing a web-based introductory BI course for non bioinformaticians that incorporates practical exercises," *Biochem. Mol. Biol. Educ.*, vol. 46, no. 1, pp. 31-38, doi:10.1002/bmb.21086.
- [17] C. Camacho and T. Madden, (2013), Mar. 12, "Blast+ release notes". Available at: <https://www.ncbi.nlm.nih.gov/books/NBK131777/>.
- [18] B. Langmead et al., (2009), "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome," *Genome Biol.*, vol. 10, no. 3, p. R25, doi:10.1186/gb-2009-10-3-r25.
- [19] H. Li and R. Durbin, (2010), "Fast and accurate long-read alignment with Burrows-Wheeler Transform," *Bioinformatics*, vol. 26, no. 5, p. 589-595, doi:10.1093/bioinformatics/btp698.
- [20] X. Wang et al., (2022), "USAT: A bioinformatic toolkit to facilitate interpretation and comparative visualisation of tandem repeat sequences," *BMC Biol.*, vol. 23, no. 1, p. 497, doi:10.1186/s12859-022-05021-1.
- [21] B. Budowle et al., (2022), "The forensic genomics toolbox is expanding". Available at: <https://www.future-science.com/doi/full/10.2144/btn-2021-0103>, *Biotechniques*, vol. 72, no. 1, 5-7, doi:10.2144/btn-2021-0103.
- [22] A. E. Woerner et al., (2017), "Fast STR allele identification with STRait razor 3.0," *Forensic Sci. Int. Genet.*, vol. 30, pp. 18-23, doi:10.1016/j.fsigen.2017.05.008.
- [23] J. L. King et al., (2021), "STRait razor online: An enhanced user interface to facilitate interpretation of MPS data," *Forensic Sci. Int. Genet.*, vol. 52, p. 102463, doi:10.1016/j.fsigen.2021.102463.
- [24] J. Hoogenboom et al., (2017), "FDSTools: A software package for analysis of massively parallel sequencing data with the ability to recognise and correct STR stutter and other PCR or sequencing noise," *Forensic Sci. Int. Genet.*, vol. 27, pp. 27-40, 2017. doi:10.1016/j.fsigen.2016.11.007.
- [25] F. R. Wendt, (2023), Chapter 19, "Bioinformatic tools for interrogating DNA recovered from human skeletal remains", In *Forensic Genetic Approaches for Identification of Human Skeletal Remains*, A. Ambers, Ed. Academic Press, 2023, pp. 433-458.
- [26] U. Lautenschlager et al., "AllCoPol: Inferring allele co-ancestry in polyploids," *BMC Bioinformatics*, vol. 21, no. 1, pp. 441, doi:10.1186/s12859-020-03750-9.
- [27] Y. Wen et al., (2023), "Forensic biogeographical ancestry inference: Recent insights and current trends," *Genes Genomics*, vol. 45, no. 10, pp. 1229-1238, doi:10.1007/s13258-023-01387-5.
- [28] R. Goel et al., (2022), "Forens (om) ic medicine: Omics in forensic medicine," *Int. J. Res. Med. Sci.*, vol. 10, no. 2, p. 557, doi:10.18203/2320-6012.ijrms20220309.
- [29] M. P. Marco-Francia and M. Mora-Ortiz (2023), "Advancements in multi-omics strategies in personalised medicine and the ethical, legal and criminological considerations," *Genoma*, vol. 58, no. 1134-7198, p. 163-182, doi:10.14679/2290.

- [30] S. Satish et al., (2023), "Future aspects of modern forensic tools and devices," *Mod. Forensic Tool. Devices Trends Crim. Investig.*, pp. 393-413.
- [31] A. A. Singh and M. Okpeku, (2024), "Emerging methods of human microbiome analysis and its forensic applications: Reviews," *Forensic Sci. Int. Rep.*, vol. 9, p. 100355, doi:10.1016/j.fsir.2024.100355.
- [32] J. Zhang, et al., (2023), "Application of microbiome in forensics," *Genomics, Proteomics and Bioinformatics*, vol 21, no 1, pp.97-107.
- [33] R. Shadoff et al., (2022), "Microbial forensics: A present to future perspective on genomic targets, bioinformatic challenges, and applications," *Forensic Genomics*, vol. 2, no. 2, pp. 42-64, doi:10.1089/forensic.2022.0007.
- [34] D. H. Huson, et al., (2016), "MEGAN community edition—Interactive exploration and analysis of large-scale microbiome sequencing data," *PLOS Comput. Biol.*, T. Poisot, Ed., vol. 12, no. 6, p. e1004957, doi:10.1371/journal.pcbi.1004957. PMID:27327495.
- [35] J. M. Robinson et al., (2020), "Forensic applications of microbiomics: A review," *Front. Microbiol.*, vol. 11, p. 608101, doi:10.3389/fmicb.2020.608101.
- [36] FK qizi Kamoliddinova and NZ. Jahongir o'g'li, (2023), "Features of the interpretation of the results of DNA examination Research,". *Educational Research in Universal Sciences*, 2(4), pp.615-620.
- [37] J. H. Smith, (2023), An exploration of the identification and processing of forensic investigative leads in investigating crime in the South African Police Service, ([Doctoral dissertation]. University of South Africa).
- [38] J. H. Smith and M. Singh, (2024), "DNA forensic and forensic investigative leads,"*J. Forensic Med.*, vol. 9, p. 347.
- [39] S. Doyle, (2024), "QHFSS DNA laboratory—ISO/IEC 17025 conformance and accreditation," *Forensic Science International: Synergy*, vol. 100449, Jan.
- [40] S. Nteziriyayo et al.,(2024), "Advancement and the existing landscape of forensic medicine in Africa: A comparison with developed countries," *Forensic Science, Medicine and Pathology*, pp. 1-14.
- [41] B. J. Gestring, (2023), "Creating infrastructure and incentives to increase quality in forensic science," *Forensic Science International: Synergy*, vol. 100435, Aug. 2023.
- [42] L. Luparia, G. Gennari, and L. Salvaderi, (2024), "DNA forensic data bank, facts and perception: The Italian experience," *Forensic Science International*, vol. 354, p. 111907, Jan.
- [43] W. Neuteboom et al., (2023), "Quality Management in forensic science: A closer inspection," *Forensic Science International*, vol. 111779, Jul.
- [44] S. Silva, (2024), "Validation of NGS for casework at forensic DNA laboratories," in *Next Generation Sequencing (NGS) Technology in DNA Analysis*, Academic Press, 2024, pp. 485-509.
- [45] J. H. Smith and J. S. Horne, (2023), "Die belangrikheid van gehaltebestuurstelsels in forensiese DNS-toetslaboratoria in Suid-Afrika," ("The importance of peer review with respect to the quality management systems in forensic DNA-testing laboratories in South Africa,") *Lit Net Akad.*, vol. 20, no. 3, pp. 491-522, doi:10.56273/1995-5928/2023/j20n3b12.
- [46] D. Stepaniuk, (2023), "Organisational Principles Of The Use Of Forensic DNA Analysis In The Investigation Of Criminal Offenses," *Archives of Criminology and Forensic Sciences*, vol. 7, no. 1, pp. 158-165, Jun.
- [47] R. E. Tontarski et al., (2024), "Alternative Models Promote Self-Regulation of the Forensic Enterprise1," in *Forensic Science Laboratory Benchmarking*, CRC Press, pp. 16-32.
- [48] R. Wickenheiser and C. Knutson, (2024), "Lead with Speed: Recasting the Forensic Laboratory Product Line and Service Delivery Model," Available at SSRN 4745595.
- [49] D. Wilson-Kovacs and D. Wyatt, (2024), "The long journey of resistance toward acceptance: Understanding digital forensic accreditation in England and Wales from a social science perspective," *Wiley Interdisciplinary Reviews: Forensic Science*, vol. 6, no. 1, p. e1501, Jan.
- [50] H.-C. Yi et al., (2022), "Graph representation learning in bioinformatics: Trends, methods and applications," *Brief. Bioinform.*, vol. 23, no. 1, Jan., p. bbab340, doi:10.1093/bib/bbab340.
- [51] S. A. Messaoudi, (2024), "Applications of NGS in analysis of challenging samples," in *Next Generation Sequencing (NGS) Technology in DNA Analysis*, Academic Press, pp. 387-411.
- [52] T. O. Olorunsogo et al., (2024), "Bioinformatics and personalised medicine in the US: A comprehensive review: Scrutinising the advancements in genomics and their potential to

- revolutionise healthcare delivery," *World Journal of Advanced Research and Reviews*, Vol 21, no.1, pp. 335–351, doi:10.30574/WJARR.2024.21.1.0016.
- [53] P. K. Yadav and S. Sharma, (2024), "Challenges in using genetic genealogy in forensics," in *Next Generation Sequencing (NGS) Technology in DNA Analysis*, Academic Press, pp. 511-530
- [54] A.H. Wheeb, A. H. (2017), "Performance analysis of VoIP in wireless networks. *International Journal of Computer Networks and Wireless Communications (IJCNWC)*, vol. 7, no. 4, pp.1-5. 2
- [55] D. A. Ajalkar et al., (2024), "Ethical and Legal Considerations in Machine Learning: Promoting Responsible Data Use in Bioinformatics," in *Applying Machine Learning Techniques to Bioinformatics: Few-Shot and Zero-Shot Methods*, IGI Global, pp. 62-74.
- [56] J. M. Butler, (2023), "Recent advances in forensic biology and forensic DNA typing: INTERPOL review 2019–2022," *Forensic Science International: Synergy*, vol. 6, p. 100311, Jan.
- [57] F. Khan and A. Mer, (2024), "The Ethical Considerations of DNA Profiling for Resilience in a Forensic Setting in India: A Comparative Study with International Guidelines," in *The Framework for Resilient Industry: A Holistic Approach for Developing Economies*, Emerald Publishing Limited, Mar, pp. 121-134.
- [58] H. Machado et al., (2024), "Ethical, legal, and social challenges of next-generation sequencing technologies (NGS) in forensic criminal identification," in *Next Generation Sequencing (NGS) Technology in DNA Analysis*, Academic Press, pp. 551-569.
- [59] S. Namasudra, (2024), "Data access control in the cloud computing environment for bioinformatics," in *Research Anthology on Bioinformatics, Genomics, and Computational Biology*, IGI Global, pp. 553-565.
- [60] M. Nogel, (2024), "Some areas where digital forensics can support the addressing of legal challenges linked to forensic genetic genealogy," *Forensic Science International: Digital Investigation*, vol. 49, p. 301696, Jun.
- [61] J.H. Smith and J.S. Horne, (2023), "Data privacy and DNA data," *IASSIST Quarterly*, vol. 47, no. 3-4, pp. 1-3.
- [62] J. H. Smith and M. Singh, (2024), "Forensic DNA Profiling: Legal and Ethical Considerations," *Journal of Scientific Research and Reports*, vol. 30, no. 5, pp. 141-144, Mar.
- [63] P. Tiwari, (2024), "Legal and Ethical Considerations in the Use of DNA Fingerprinting," *Journal of Scientific Research and Reports*, vol. 30, no. 3, pp. 236-242, Feb.
- [64] P. K. Yadav and S. Sharma, (2024), "Challenges in using genetic genealogy in forensics," in *Next Generation Sequencing (NGS) Technology in DNA Analysis*, Academic Press, pp. 511-530.

AUTHORS

Joe H Smith, The author is responsible for managing and administrating the National Forensic DNA Database of South Africa. Having accumulated more than 36 years of experience in forensic services, he has significantly contributed to developing a law that governs the collection and processing of fingerprints and DNA. Furthermore, he spearheaded the development of an integrated system solution for managing a forensic DNA database. The author possesses a master's degree in technology and is pursuing a doctoral degree in philosophy, focusing on criminal justice, police science, and forensic science. The author has written and collaborated on several international and national academic journal publications. In addition, he has made contributions to chapters in books in the field of forensic science and investigation.



M Singh, The author is an experienced biotechnologist and DNA forensic analyst, boasting 26 years of combined expertise in animal research, healthcare, and forensic DNA and data analysis.

