International Journal of Biological, Physical and Chemical Studies
ISSN: 2709-1554
DOI: 10.32996/ijbpcs
Journal Homepage: www.al-kindipublisher.com/index.php/ijbpcs

IJBPCS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Synergizing Big Data and Biotechnology for Innovation in Healthcare, Pharmaceutical Development, and Fungal Research

Ahmed Tanvir[1], Bismi Jatil Alia Juie[2], Nadia Terasa Tisha[3], Md Majedur Rahman[4]✉
[1]*College of Pharmacy, Kyungpook National University, Daegu 41566, Korea*
[2]*Training Department, Incepta Pharmaceuticals Limited, Dhaka-1208, Bangladesh*
[3]*Faculty of Business Administration, Dhaka University, Dhaka-1000, Bangladesh*
[3]*Medical Services Department, Incepta Pharmaceuticals Limited, Dhaka-1208, Bangladesh.*
Corresponding Author: Md Majedur Rahman, **E-mail**: sujan.smc@gmail.com

| ABSTRACT
The convergence of big data and biotechnology is transforming the landscape of healthcare, pharmaceutical research, and fungal biology. This review explores the emerging synergy across these domains, emphasizing predictive analytics, artificial intelligence (AI), and machine learning (ML) enabling real-time decision-making, accelerating drug discovery, and advancing ecological and mycological research. In healthcare, big data collected from electronic health records (EHRs), wearable devices, and population-level datasets support early disease detection, risk stratification, and personalized treatment plans. In pharmaceuticals, AI models including deep learning and generative framework streamline drug development by facilitating target identification, virtual screening, and predictive ADMET modelling. These innovations have significantly reduced development timelines and improved precision in therapeutic design. Parallel advancements in fungal biotechnology, driven by image-based classification and genomic analysis, are revealing fungi as critical sources of bioactive compounds, enzymes, and ecological indicators. Predictive models are now capable of identifying fungal species, mapping metabolic pathways, and forecasting ecological patterns, thus positioning fungi at the intersection of environmental monitoring and drug discovery. Despite these advances, challenges persist including data interoperability, algorithmic bias, regulatory barriers, and ethical concerns related to privacy, equity, and bioprospecting. This review also discusses the infrastructure needed to support cross-sector innovation, such as cloud computing, graph neural networks, FAIR data standards, and open science platforms. It outlines strategic priorities for building integrated, explainable, and accessible AI systems, particularly in underserved regions. By highlighting case studies, shared challenges, and future directions, the review underscores the importance of interdisciplinary collaboration in leveraging big data–biotech synergy.

| KEYWORDS
Big Data Analytics, Biotechnology, Predictive Modelling, Fungal Bioinformatics

| ARTICLE INFORMATION

ACCEPTED: 10 December 2020 **PUBLISHED:** 28 December 2020 **DOI:** 10.32996/ijbpcs.2020.2.2.4

1.0 Introduction

Over the course of the past few years, the convergence of big data and biotechnology has brought about a profound shift in the way that we approach health services, medications, and the biological sciences. Big data, which encompasses electronic health records, wearable sensor streams, genetic sequences, chemical libraries, and environmental observations, presents both enormous opportunities and significant analytical obstacles (Adans-Dester et al., 2020; Abdullahi, 2011). Big data encompasses massive datasets that are complex in nature. In the meantime, improvements in biotechnology in the areas of computational modeling, high-throughput screening, and molecular engineering have resulted in the development of strong tools that are able to extract insights that can be put into practice from this data.

These sectors, when combined, make strides in customized medicine, speed up the development of new drugs, and expand into developing fields such as fungal biotechnology, which enables advancements in a variety of areas, ranging from clinical diagnostics to ecological monitoring (Bulbul et al., 2018; Chaudhary & Khadabadi, 2012). Big data is a term used in the fields of biology and medicine to describe datasets that are so extensive or intricate that they beyond the capabilities of conventional processing methods as determined by the "4 Vs": volume, variety, velocity, and veracity. The following categories of data are included in this category: organized sources, which include patient medical records and the findings of chemical screening ; semi-structured and unstructured sources, which include imaging, time-series sensor logs, and genomics; and unstructured sources. It is necessary to utilize advanced analytics in order to effectively manage such a vast amount of data (Ma et al., 2020; Manik et al., 2018). These analytics include machine learning (ML), deep learning (DL), natural language processing, and graph-based models. These analytical methods are transformed into solutions that can be implemented through the use of biotechnology. In the field of medicine, real-time clinical choices are informed by diagnostics and detection technologies that are powered by machine learning. Target identification and ADMET (absorption, distribution, metabolism, excretion, and toxicity) prediction are only two examples of how predictive models and generative artificial intelligence are reshaping the drug pipeline in the pharmaceutical industry (Jonathan et al., 2020). Researchers are discovering new enzymes, medicinal chemicals, and ecological markers through the use of image analytics, genomics, and metabolomic screening in the field of fungal exploration. When taken as a whole, this synergy represents a transition away from compartmentalized approaches and toward integrated, data-driven life sciences (Rosa et al., 2019; Rubina et al., 2017; Sitarek et al., 2020; Tobore et al., 2019).

Big data is the driving force behind the era of Healthcare 4.0, which is characterized by the digital integration of electronic health records (EHRs), medical imaging, wearable devices, and Internet of Things (IoT) systems (Miah et al., 2019). In order to promote early disease detection, hospital resource optimization, and tailored treatment planning, predictive models that make use of various data streams are an invaluable tool. Systems that make use of machine learning and cloud computing have made it possible for predictive algorithms to identify cardiovascular events or acute care requirements (arxiv.org). Data from electronic health records (EHRs) and wearables have shown promise in risk forecasting, despite the fact that significant clinical integration is still in its infancy due to difficulties in data governance, workflow compatibility, and reliability. Big data applications were expedited as a result of the COVID-19 pandemic, which led to the development of real-time case modeling, supply chain forecasting, and AI-guided therapeutic exploration. These applications were essential in the management of global health emergencies.

2.0 Big Data Applications in Healthcare

Wearable sensors, predictive diagnostics, electronic health record analytics, and integrated artificial intelligence interventions are driving the transition to a new era of dynamic, data-driven medicine, which is being ushered in by the ongoing confluence of big data and healthcare. In order to improve health outcomes, optimize workflows, and provide support for clinical decision-making, each of these domains makes use of large-scale and multimodal data (Miah et al., 2019; Andrew et al., 2020; Arpaia et al., 2013).

2.1 Real-Time Health Monitoring with Wearables

Some of the physiological signals that can be continually captured by wearable devices include heart rate, heart rate variability, body temperature, and accelerometry. These signals can be collected by smartwatches, electrocardiogram patches, and biometric patches, among others. These data streams are fed into machine learning (ML) models that are aimed to identify early warning symptoms of cardiovascular disease (CVD) and other health risks (Andrew et al., 2020; Aerts, 2020; Allegra, 2019). According to a comprehensive assessment of 55 research, machine learning models that were trained on wearable data were able to accurately predict the outcomes of cardiovascular disease (CVD). However, none of these models were able to reach clinical deployment (TRL < 6). Furthermore, none of these models contained prospective phase 2/3 trials, and the majority of them required external validation (Aerts, 2020). Ensemble models such as Random Forest and deep architectures obtained above 90% accuracy in diagnosing hypertension, arrhythmia, and heart failure, according to the findings of another systematic investigation that focused on Internet of Things-based cardiovascular monitoring. This analysis included 164 papers. In addition, CNN–LSTM and hybrid neural network architectures have been successfully implemented on microcontroller units (such as the Cortex M4) in order to execute electrocardiogram (ECG) classification procedures with F1 scores that are greater than 0.78 (Dongmei et al., 2020). Even if the performance is great, there are still significant gaps that exist. These gaps include inconsistent sampling rates, sensor heterogeneity, annotation variability, and hardware limits that hinder translational attempts. There are still very few devices that have been approved by the FDA for cardiovascular disease monitoring, and obstacles like as multi-sensor data fusion, battery life, and privacy concerns continue to slow down progress.

2.2 Predictive Diagnostics for Chronic Disease

Aside from wearables, big data is the driving force behind predictive diagnoses for chronic disorders such as diabetes, hypertension, and heart disease. This is accomplished through tailored risk modeling methods. Early disease identification and risk stratification are both made possible by the use of longitudinal electronic health records (EHRs) in conjunction with machine learning (Dongmei et al., 2020). According to the findings of a scoping review of machine learning applied to longitudinal electronic health records (EHRs), there is widespread application across chronic disease domains. Models have demonstrated their usefulness

in early diagnosis, prevention, and risk modeling. However, rigorous evaluations of clinical utility rarely exist. Lee et al., 2020 and Meyer et al., 2020 presented the results of another study that demonstrated the use of deep representation learning to more than 1.6 million electronic health records (EHRs) of patients. The researchers used autoencoders and embeddings to create latent features that successfully classified illness subtypes in neurological disorders and type 2 diabetes. The use of predictive analytics in personalized medicine helps to bridge the gap between patient phenotypes and genetics. EHR Platforms that have been upgraded with artificial intelligence have been utilized, for example, to match phenotypic and genomic data, which has resulted in the quicker recognition of hereditary problems in neonates. According to Lee et al. (2020) research, federated learning involves multiple institutions and enhances the generalizability of models while also protecting patient privacy in uncommon disease prediction. Despite this, electronic health record data continues to be disorganized, as it is fragmented, heterogeneous, and biased. When used in predictive models, it frequently necessitates extensive preparation. Information that is both structured and unstructured (Manik, 2020).

2.3 EMR/EHR Analytics & Personalized Medicine

Large-scale deep learning models that ingest raw electronic health records (EHR) in the format of Fast Healthcare Interoperability Resources (FHIR) have demonstrated superior performance in predicting death (AUC ~0.93–0.94), readmission (AUC ~0.75–0.86), and diagnosis (AUC ~0.90) across numerous health systems, surpassing the performance of alternative risk scores. Compared to uni-modal models, multimodal fusion of medical imaging and electronic health record (EHR) data, which makes use of convolutional neural networks (CNNs) linked with structured clinical information, has also proven superiority in illness diagnosis.

In order to provide real-time clinical decision support, these platforms facilitate the matching of patient profiles with insights that may be put into action. However, there are not many rigorous prospective trials, and there are not enough standardized case studies or performance benchmarks in live clinical settings (Ma et al., 2020). Biotechnology and big data are actively altering the healthcare industry through the implementation of real-time monitoring, precision diagnostics, and decision-making that is based on evidence. Wearables and electronic health record analytics provide predictive diagnostics that have been shown to be very accurate; nevertheless, in order to translate these models into clinical tools, it is necessary to address concerns with data quality, regulatory validation, and an equal access to the information. As edge artificial intelligence, federated learning, and multimodal integration continue to develop and mature, healthcare systems are getting closer and closer to a tipping point when predictive analytics will become a standard component of everyday care delivery (Ma et al., 2020; Meyer et al., 2020).

3.0 Biotech Driven Innovations in Pharmaceuticals

The fusion of biotechnology and big data is revolutionizing pharmaceutical R&D. Cutting-edge AI models especially generative architectures, predictive ADMET systems, and genomics platforms are creating powerful, data-intensive pipelines that challenge traditional development paradigms. These approaches are transforming how drugs are discovered, optimized, and brought to clinical readiness (Allegra, 2019).

3.1 AI in Drug Discovery and Development

Graph neural networks (GNNs), transformers, Variational Autoencoders (VAEs), and Generative Adversarial Networks (GANs) are some examples of generative artificial intelligence models that have emerged as powerful tools for molecule generation, virtual screening, and lead optimization. Recent studies have shed light on the expanding role that generative artificial intelligence plays in the process of developing new drugs. These studies demonstrate how deep learning can navigate huge chemical regions in order to offer structurally innovative candidates that are effective against specific targets (Aerts, 2020; Allegra, 2019). By screening only a small portion of compound libraries, active learning techniques paired with deep docking can find 80–90% of hits, hence substantially lowering the amount of computational resources required (Sitarek et al., 2020; Tobore et al., 2019). This is something that reviewers have pointed out. It is another momentous occasion because AlphaFold 2 has achieved success in precisely predicting protein structures in the year 2020. It does this by providing high-fidelity predictions of protein folding, which enables more educated target identification and binding-site modeling, which in turn strengthens AI-driven drug discovery based on structural data (Sitarek et al., 2020). The authors warn that there are still challenges to general adoption, despite the fact that the potential is enormous. These obstacles include the explainability of algorithms, biases in training data, and a need for consistent standards.

3.2 Big Data Platforms in Genomics and Pharmacogenomics

The incorporation of genomic and pharmacogenomic databases into drug development platforms offers increased precision and individualization in the design of therapeutic interventions. There is the potential for multi-omic data to inform pharmacological mechanisms and patient stratification when large biobank–EHR-linked cohorts are utilized. One of the benefits that has been highlighted in rheumatologic precision medicine studies is that AI-based pharmacogenomics systems have the ability to assess patient genotypes in real time, which then allows for the optimization of dose or drug choice (Lee et al., 2020; Meyer et al., 2020). In order to facilitate clinical application through CPIC guidelines all over the world, PharmGKB, which is a curated pharmacogenomic knowledgebase, provides support for this integration by correlating genomic variations to medication

responses at the genomic level. These systems are able to examine data from several institutions while maintaining confidentiality as artificial intelligence-powered platforms and federated learning models become more prevalent. Despite the progress that has been made, the diversity and accessibility of large-scale genomic datasets are not uniform, which makes it difficult to train and validate models in a manner that is equitable across populations (Abdullahi, 2011; Manik, 2020).

3.3 Predictive Models for ADMET and Target Validation

Maintaining an early prediction of therapeutic efficacy and safety is of utmost importance. Artificial intelligence-driven ADMET platforms that make use of multi-task deep learning have exhibited significant advancements, outperforming traditional QSAR methods in some cases. Improved lead selection workflows can be achieved through the utilization of tools like as ADMET AI, which offer quick toxicity and metabolism predictions on demand (Abdullahi, 2011). Target validation can also be improved by using knowledge-graph and artificial intelligence technologies. These methods make use of literature mining and data from biomedical networks to propose new targets and repurposed medications. Explainable models that provide toxicity-associated molecular substructures or ADMET explanations are crucial to establish trust in these systems; nonetheless, openness in these systems is essential.

3.4 Strategic Models for Reducing Development Timelines

Artificial intelligence-driven firms have started delivering lead compounds at timelines that have never been seen before. In less than fifty days, Insilico Medicine was able to offer potential therapeutic leads by merging generative chemistry, target scoring, and ADMET modeling through the usage of its PandaOmics and Chemistry42 artificial intelligence systems. According to Abdullahi (2011), XtalPi and Pfizer both utilized quantum-informed artificial intelligence in order to speed up the process of identifying Paxlovid components. Companies such as GSK and Exscientia are reportedly working on AI-designed preclinical candidates that are making their way into clinical trials, according to Wired and other sources. Nevertheless, sustained success is not solely dependent on the speed of artificial intelligence; it also depends on comprehensive validation processes, which include prospective assays, safety studies, and regulatory alignment.

3.5 Integrated Pipelines & End-to-End Systems

The development of end-to-end artificial intelligence drug pipelines that include generative chemistry, structure prediction, synthesis planning, ADMET filtering, and clinical simulation is the most significant trend that has emerged in recent years. Platforms such as AMPL (ATOM consortium) are examples of integrated workflows that provide candidate sets that are optimized across several criteria with minimum interaction from humans. This integration is improved by pipelines that are enabled with AlphaFold. These pipelines combine structure-based screening with generative chemistry, which enables on-the-fly optimization of candidate binding. Rose et al. (2019), Sitarek et al. (2020), and Tobore et al. (2019) have all pointed out that in order to realize these platforms at scale, they require standardized data ecosystems, modular toolkits, and explainability that is ready for compliance. This presents a clear road toward comprehensive AI-first medication research and development. Pharmaceutical research and development is being reshaped by the synergy between big data and biotechnology. Drug development is being revolutionized by a combination of factors, including faster pipelines, genomics-based platforms, ADMET prediction frameworks, and generative models powered by artificial intelligence. Emerging frameworks such as self-supervised learning, federated models, and hybrid mechanistic integrations offer a realistic road ahead, despite the fact that issues in data governance, interpretability, and legislation continue to exist. With the deployment of end-to-end artificial intelligence systems by both startups and pharma giants, the future holds a great deal of promise: drug discovery that is quicker, less expensive, and more accurate, all of which will be driven by intelligent data ecosystems (Manik, 2020; Adans-Dester et al., 2020).

4.0 Predictive Analytics and Big Data in Mycology

Advancements in machine learning (ML), deep learning (DL), and big data analytics are rapidly transforming mycology from species identification and ecological forecasting to pharmaceutical discovery and disease modeling (Adans-Dester et al., 2020; Ma et al., 2020; Manik et al., 2018). Secondary Metabolism Mining" emphasizes the exploration of fungal metabolites for pharmaceuticals. "Ecological Monitoring & Fungal Disease" highlights the role of remote sensing and AI in ecosystem and disease surveillance. "AI-Guided Bioactivity Screening" enables prioritization of potent bioactive compounds using machine learning, while "Functional Genomics & Expression" leverages genomic data to enhance fungal biosynthetic pathways. Finally, "Disease Modeling Platforms" use predictive tools to identify pathogenic fungi and model their interactions. Together, these interconnected approaches advance diagnostics, drug discovery, and fungal ecology through data-driven innovation (Figure 1).

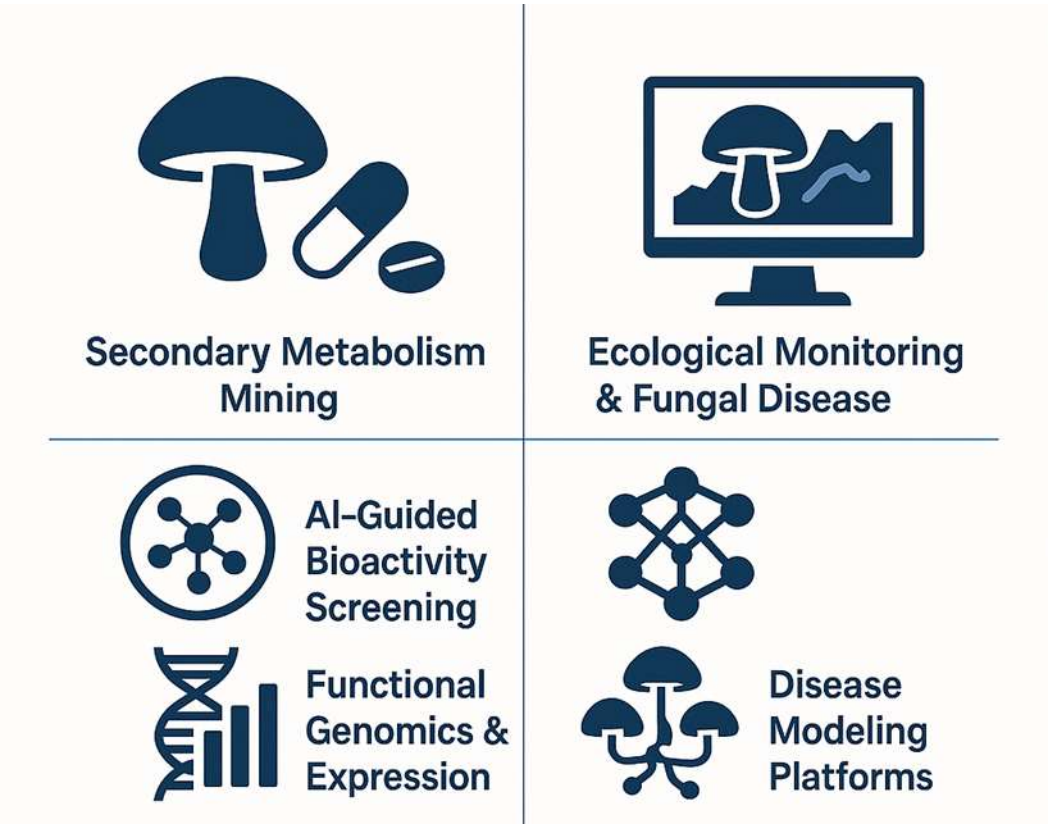

Figure 1. Predictive analytics and big data in mycology.

4.1 Fungi in Biopharma: Antibiotics, Enzymes, and Bioactive Molecules

Fungi's evolutionary diversity gives rise to rich bioactive molecule repertoires ideal for pharmaceuticals and industrial enzymes in previous research (Lee et al., 2020; Meyer et al., 2020):

• Secondary metabolism mining: Analysis shows many fungal SMs are more "drug-like" than bacterial counterparts under FDA metrics. This supports their continued exploration for antibiotics, immunomodulators, and anticancer agents.

• AI-guided bioactivity screening: Predictive models trained on combined bacterial and fungal BGCs help prioritize high-value fungal metabolites, especially urgent given antibiotic resistance.

• Functional genomics & expression: Genomics-informed ML reveals biosynthetic pathway activation patterns, enabling metabolic engineering to enhance production of target bioactive compounds potentially transforming industrial biotech research.

• Disease modeling platforms: Data-informed fungal models help identify virulent fungi and potent metabolites while also informing fungal pathogen–host interaction frameworks.

4.2 Ecological Monitoring & Fungal Disease Modeling

Big data supports fungal ecology from remote sensing to predictive disease alerts:

• Hyperspectral remote sensing: Self-supervised deep DL combined with hyperspectral imagery identified Fusarium head blight in wheat, enabling early asymptomatic detection.

• Mycetoma histopathology: The MyData database, featuring 864 labeled histopathological images in mycetoma cases, supports AI-based screening of tissue samples and an important step toward automated diagnostics in endemic areas.

• Ecosystem forecasting: Models using Landsat and satellite indices have forecasted fruiting patterns and health of macrofungi (e.g., Lactarius deliciosus), providing tools for forest management.

• Fungal-fungal interaction networks: DL pipelines now classify fungal-fungal interactions enabling prediction of species competition, invasive risks, or synergistic community shifts.

Mycological research is extending into areas that were previously dominated by human competence thanks to the developments of predictive analytics and big data. Fungal science is approaching a new data-driven frontier, which includes the identification of species in less than twenty-four hours by time-lapse imaging and the development of antibiotics guided by BGC (Rosa et al., 2019; Rubina et al., 2017; Das et al., 2016, 2017). The integration of imaging, genetic, metabolic, and ecological data holds the promise of establishing strong pipelines that can support the activities of biodiversity monitoring, illness diagnostics, and the creation of biopharmaceuticals. The maturation of prediction tools from proof-of-concept to effective applications in both research and

society will be ensured by the continued development of multi-modal datasets, hybrid model transparency, and field validation frameworks.

5.0 Data Integration, Infrastructure, and Tools

The entire potential of the synergy between big data and biotechnology can only be realized through the utilization of a solid infrastructure, seamless interoperability, and shared tools. Based on Rosa et al.'s 2019 research, this section examines the most important platforms, standards, and frameworks that facilitate the integration of data from different domains (such as the cloud, Hadoop, and GNNs), handle difficulties related to format and metadata, and embrace the concepts of FAIR and open-science.

5.1 Platforms, Pipelines, and Analytics Infrastructure

Big-data frameworks (Hadoop, Spark) and cloud-based platforms (AWS, Azure, and Google Cloud) have become indispensable in the fields of environmental biology, healthcare, and research and development in the pharmaceutical industry. Cloud architecture allows scalable data processing, safe data sharing, and machine learning-driven analysis across any industry, according to a review that was published not too long ago. Manik (2020) stated that these settings are capable of supporting high-performance storage, distributed compute, and elastic scaling features. Toolkits such as Nextflow and Galaxy stand out as particularly useful for bioinformatics workflows. Nextflow offers a portable and robust engine that can execute pipelines in a variety of contexts, including local, high-performance computing, and cloud environments. Additionally, it supports containerization and community-driven standards.

5.2 Graph Neural Networks & Knowledge Graphs

The foundation of complex relationships in the fields of healthcare (electronic health record networks), drug research (molecular interactions), and fungal ecology (species interactions) is provided by graph-structured data. An analysis of the uses of GNNs in the medical field reveals that their strengths lie in the prediction of diseases and the identification of new drugs, as well as in the modeling of relationships through the use of attention mechanisms and graph convolution. According to the findings of another assessment in the field of computer vision, GNNs are suitable for multimodal data fusion, which is a requirement that becomes particularly pressing when merging imaging, genomic, and ecological data (Allegra, 2019).

5.3 Interoperability and Data Standards

Fast Healthcare Interoperability Resources (FHIR) is the foundation upon which interoperability in the healthcare industry is built. FHIR defines modular, RESTful data formats for electronic health record (EHR) sharing. Public implementation of FHIR in systems all across the world, including mandates from the United States Centers for Medicare and Medicaid Services (CMS) and programs in Brazil and Israel, improves real-time data sharing and clinical pipeline alignment. According to Ma et al. (2020) and Manik et al. (2018), data pipelines that are compatible with FHIR make it possible to perform federated machine learning in the cloud without having to centralize sensitive health records.

5.4 Open Science & FAIR Data Practices

It is possible to ensure that data generated across disciplines becomes a common resource by adhering to the FAIR principles, which stand for "findable, accessible, interoperable, and reusable". An example of this would be a study that was conducted by the BMC that underlined the significance of incorporating FAIR into training programs in order to enhance the repeatability and literacy of researchers (Andrew et al., 2020; Allegra, 2019). Reviews in the biopharmaceutical industry call for the adoption of FAIR data in order to significantly improve collaboration and data value. Additionally, FAIR frameworks have been effectively implemented in the fields of genomic monitoring, global health, and mental health in order to guarantee the ethical reuse of shared resources. The following are examples of tools that are used to promote FAIR practices: repository platforms such as Zenodo and Figshare; BioRxiv and medRxiv for preprint sharing and metadata; Galaxy and nf-core for embedding metadata and version control by design. Transparency, reproducibility, open peer review, and equal access are all encouraged by these infrastructures. Interoperability, infrastructure, and FAIR data are the three pillars that support the synergy between big data and biotechnology. The use of integrated technologies enables innovation in a variety of fields, including cloud-native pipelines in the healthcare industry, standardized genomic workflows in the pharmaceutical industry, and open mycology databases. Within the realm of data-driven science, the establishment of shared standards, capacities, and open frameworks is absolutely necessary for achieving sustainable and equitable progress.

6.0 Challenges and Ethical Considerations

6.1 Data Privacy and Security Across Sectors

In healthcare and biotech, data is governed by frameworks like HIPAA (U.S.) and GDPR (EU), but AI systems often exploit subtle gaps. Commercial AI implementations can expose sensitive medical data via de-anonymization or insecure storage on third-party platforms. Medical robotics and wearable sensors further expand the attack surface, potentially enabling malicious access to private health data. Precision health models integrating EHR, genomic, and device streams require advanced privacy-preserving

cryptographic techniques such as federated learning and homomorphic encryption. Establishing secure environments for cross-institutional AI development while maintaining patient consent and trust is a central necessity.

6.2 Bias in Training Data and Model Generalizability

Artificial intelligence models reflect and amplify the biases that are present in the training data, particularly when the datasets are not representative of the populations that they serve. This results in biased consequences in the healthcare industry with regard to racial bias, gender bias, ethnic bias, and socioeconomic level barriers. An example that is particularly noteworthy is a clinical algorithm that uses the cost of healthcare as a proxy for patients who are systematically under-selected who are black. In global deployment, models that are adjusted in high-income settings frequently fail in low and moderate income contexts (for example, hospitals in the United Kingdom versus hospitals in Vietnam). Fairness audits, diversified benchmarking datasets, and data partitioning algorithms that are reinforced by transfer learning and threshold calibration are some of the emerging techniques that are being developed to fight this particular issue. In order to prevent the over-representation of populations or taxa that have been thoroughly researched, it is necessary to balance the bias that exists in biotech data, notably in pharmacogenomic and fungal species databases.

6.3 Regulatory Hurdles in AI-Enabled Biotech Systems

Artificial intelligence-driven solutions that continually learn or operate across modalities are not well matched to the regulatory frameworks that are already in place. With scant guidelines on how to evaluate changing models, the Food and Drug Administration (FDA) and other authorities continue to treat algorithmic systems as either fixed software medical devices or emergent software medical devices (SaMD). There is a growing trend towards mandating model interpretability and transparency; yet, a significant number of generative or deep neural network models continue to be opaque (Sitarek et al., 2020; Tobore et al., 2019). The regulatory acceptance process is further complicated by concerns regarding the quality of the data, particularly in genomics or ADMET prediction. However, despite its significance, the harmonization of pipelines through the use of standards such as TRIPOD-AI, DECIDE-AI, BioCompute, and CWL is frequently neglected. Additional standards are required for the utilization of genetic resources, the provenance of samples, and the protection of biosafety in the context of fungiculture and bioprospecting (Sitarek et al., 2020).

6.4 Ethical Concerns in Fungal Bioprospecting and Health Surveillance

In the process of bioprospecting fungal biodiversity, environmental, social, and legal challenges are brought to light. Despite the fact that fungus have a wealth of antibiotics and enzymes, harvesting them without proper regulation can be detrimental to ecosystems and infringe the rights of communities. An example of the dangers that can arise from failing to obtain community consent or benefit-sharing is the Maya ICBG case that occurred in 1999. At this point in time, ethical frameworks place an emphasis on prior informed permission, equitable benefit-sharing, and conservation measures. Fungal monitoring technologies, particularly those that involve genomic sequencing, are careful to adhere to privacy standards when it comes to environmental sampling when it comes to public health surveillance. Because the data may incriminate communities in situations where fungal exposures map to human activities, it is imperative that sampling and utilization be conducted in an open and honest manner (Sitarek et al., 2020). Big data and biotechnology working together can produce prospects that are transformative, but only if they are managed with a high level of ethical care. In order to provide sustainable healthcare, ethical utilization of fungal resources, and equitable pharmaceutical innovation, artificial intelligence that also protects privacy, is impartial, is sympathetic to regulations, and is socially responsible is required. Future advancement is contingent upon governance that is open and transparent, the development of global capacity, and frameworks that are inclusive and that respect both innovation and integrity.

7.0 Future Directions and Cross-Sector Synergy

The figure illustrates a forward-looking vision for cross-sector integration of biotechnology and big data, centered around an "Integrated Bio-Intelligence Ecosystem." This ecosystem connects four critical domains that together reshape fungal research, healthcare, and pharmaceutical innovation. Integrated health–pharma–fungal platforms leverage EHRs, genomics, wearable technologies, and fungal metabolites under a unified One Health approach. Hybrid biotech–data pipelines transform fungal biosynthetic gene cluster (BGC) data into actionable drug leads using AI-based prediction, cheminformatics, and iterative lab validation. Scalable AI systems for remote populations promote global health equity through federated learning, edge deployment, and cloud tools. Finally, strategic policy and investment drive the ecosystem forward through FAIR data mandates, cross-sector funding, bioconvergence hubs, and regulatory standards for explainable and safe AI. Collectively, this integrated framework enables holistic, data-driven solutions to complex biological and public health challenges (Figure 2). As biotechnology continues to intersect with big data, the path forward lies in building holistic, integrated platforms that bridge healthcare, pharma, and fungal research. These systems will enhance therapeutic discovery, public health, and ecosystem monitoring, especially in under-resourced settings. Four key future directions emerge:

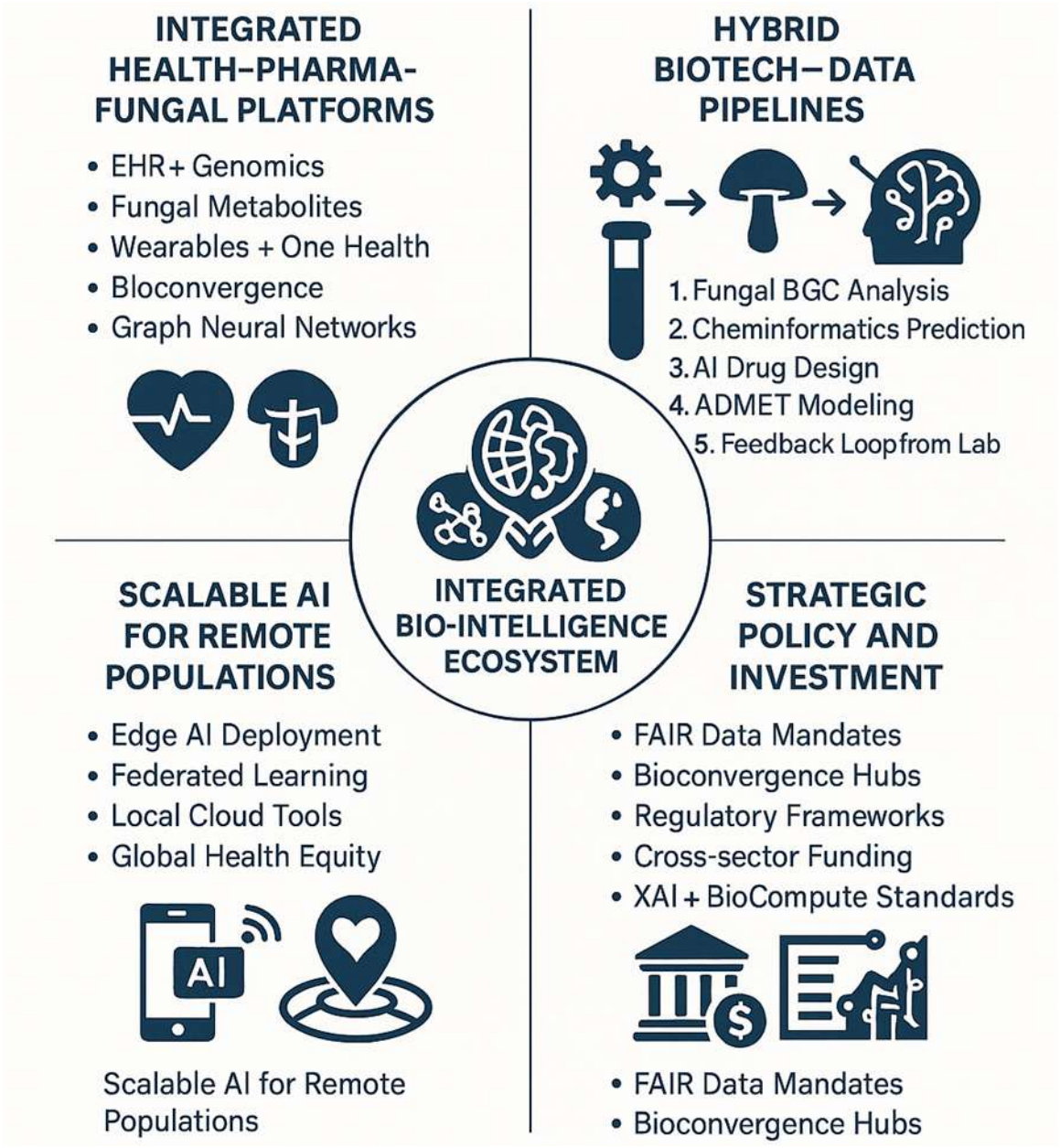

Figure 2. Converging data and biology to reshape health, pharma, and ecology.

7.1 Toward Integrated Platforms Combining Health, Pharma, and Fungal Intelligence

The concept of bioconvergence bringing together diverse biological and data-driven disciplines is gaining momentum. Future platforms could combine EHRs, wearable data, genomic/pharmacogenomic insights, chemical libraries, fungal metabolite profiles, and environmental metadata to create unified biosystems (Sitarek et al., 2020; Tobore et al., 2019; Ma et al., 2020; Manik et al., 2018 ; Aminuzzaman & Das, 2017; Marzana et al., 2018).

• In drug development, insights drawn from fungal biosynthetic gene clusters (BGCs) and macro-molecular ecological signals could be integrated with patient-specific pharmacogenomic data to speed up bioactive compound discovery.

• In public health, wearable physiological data could trigger environmental fungal surveillance alerts bolstering early warnings under One Health initiatives.

These platforms can be powered by hybrid AI pipelines leveraging graph neural networks, federated learning, and provenance tracking (via BioCompute/RO-Crate frameworks) delivering transparent, cross-domain analytics at scale.

7.2 Potential of Hybrid Biotech–Data Pipelines

A critical frontier lies in workflows that transform fungal metabolite data into drug lead candidates:

1. Genomic and metabolomic analysis of fungi to identify promising BGCs.
2. Predictive cheminformatics to forecast novel compound activities (e.g., AI for antimicrobial peptides).
3. Generative AI design to create optimized analogs tailored to human pharmacokinetics and safety.
4. Virtual screening and ADMET modeling to assess efficacy and toxicity before lab testing.
5. Closed-loop lab validation, feeding results back into AI systems for continuous improvement.

This hybrid ecosystem spanning omics, AI, simulation, and bioprospecting offers a pipeline for rapid, data-driven natural product drug discovery with impact on antimicrobial resistance and other public health priorities.

7.3 Scalable AI Systems for Underserved and Remote Populations

AI platforms must be accessible globally, particularly in regions with high health burden and low resources. Key strategies include:

• Edge deployment of AI on wearables, smartphones, and portable imaging devices, reducing dependence on central compute.

• Federated learning frameworks to support distributed model training without sharing sensitive data, crucial for protecting patient privacy.

• Affordable cloud resources and modular AI-as-a-service platforms enabling local institutions to analyze genomics, drug potential, fungal outbreaks, and health records, regardless of infrastructure constraints.

• Tools like federated patient monitoring and AI-powered diagnostics are already being trialed for cardiovascular care, infectious diseases, and environmental health monitoring.

These approaches democratize access to AI-driven healthcare and biosurveillance, reducing global inequities.

7.4 Strategic Policy & Funding Priorities for Biotech–Big Data Ecosystems

To support such ecosystems, forward-thinking policy and investment are essential based on the previous research (Das & Aminuzzaman 2017; Das et al., 2016).

• Incentivize open, FAIR-compliant data sharing : Mandate metadata-rich publication in public repositories (e.g., fungal BGCs, EHR phenotypes, wearable signals) and support FAIR training programs.

• Support bioconvergence hubs that co-locate disciplines across AI, biology, forestry, healthcare, and pharma similar to Israel's AION Labs ecosystem.

• Promote cross-sector translational funding: Funds should specifically prioritize projects combining fungal metabolites with drug pipelines, distributed health monitoring, or environmental health surveillance accelerating ecosystems with social impact.

• Update regulatory frameworks for AI-generated bioactives and diagnostics : Establish pre-competitive standards for explainability (e.g., XAI benchmarks), provenance (BioCompute), and trial validation of edge-AI tools.

• Invest in hybrid capacity-building : Support training programs that combine biology, computational science, ethics, and data stewardship enabling the next generation to manage high-dimensional, cross-disciplinary projects.

8. Conclusion

The combination of big data and biotechnology is altering the way in which we handle some of the most important concerns in the fields of environmental research, medicines, and health care. Real-time disease monitoring, precision drug discovery, high-throughput fungal classification, and bioprospecting are now all possible thanks to predictive analytics, which is powered by artificial intelligence and machine learning. The use of big data systems in the healthcare industry has the potential to enhance diagnoses, personalize therapy through genomics, and enable scalable remote monitoring through the utilization of wearable devices. In the field of pharmaceutical innovation, platforms driven by artificial intelligence have contributed to a reduction in lead identification delays, an improvement in drug safety through predictive ADMET modeling, and the opening of the door to the rapid creation of medications of the next generation. Image-based identification, genetic trait prediction, and integration into biopharmaceutical processes are all contributing to the exciting new developments in fungal research that are taking place simultaneously. In conjunction with one another, these cross-sectoral advancements highlight the revolutionary potential of the synergy between big data and biotechnology. This shift is made possible in large part by the introduction of frameworks that are transdisciplinary in nature. However, the most significant innovations are not the result of separate advancements in computers or biology; rather, they are the result of the planned integration of these two fields, which is becoming increasingly known as bioconvergence. These frameworks bridge the gap between previously compartmentalized fields of study, whether it be through the integration of fungal metabolite screening with artificial intelligence medication modeling or through the combination of environmental fungal surveillance with wearable health data for the purpose of public health risk assessments. The operationalization of these complex systems is currently supported by a variety of tools, including federated learning, graph neural networks, FAIR data standards, and cloud-native bioinformatics pipelines. Consequently, the development of cooperation between healthcare professionals, environmental scientists, pharmacologists, and data engineers will be essential to the achievement of future success opportunities. With an eye toward the future, it will be of utmost importance to make certain that these advances are not just technologically advanced but also equitable and sustainable. The rise of technology must be accompanied by strategic investments in infrastructure, regulatory adaption, data ethics, and capacity-building, particularly in places that are underrepresented. The advantages of this transition will be distributed more evenly around the world if open research, explainable

artificial intelligence, and inclusive policies are implemented. As we move into an era of intelligent, integrated biotechnology, the focus must continue to be on the creation of resilient systems that utilise data in an ethical and efficient manner in order to enhance health, promote discoveries, and safeguard ecosystems. The convergence of big data and biotechnology is more than just a trend in the scientific community. A structural shift that has the potential to reimagine creativity in the 21st century is being brought about.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Abdullahi, A.A. Trends and challenges of traditional medicine in Africa. Afr. J. Tradit. Complement Altern. Med. 2011, 8, 115–123.
[2] Adans-Dester C, Bamberg S, Bertacchi F, et al. Can mHealth Technology Help Mitigate the Effects of the COVID-19 Pandemic?. IEEE Open Journal of Engineering in Medicine and Biology 2020;1:243
[3] Aerts J. Special Issue on "Human Health Engineering". Applied Sciences 2020;10(2):564
[4] Allegra, M. Antioxidant and anti-inflammatory properties of plants extract. Antioxidants 2019, 8, 549.
[5] Aminuzzaman FM, Das K (2017). Morphological Characterization of Polypore Macro Fungi Associated with Dalbergia Sissoo Collected from Bogra District Under Social Forest Region of Bangladesh. Journal of Biology and Nature. 6(4):199–212.
[6] Andrew C Liu, Krishna Patel, Ramya Dhatri Vunikili, Kipp W Johnson, Fahad Abdu, Shivani Kamath Belman, Benjamin S Glicksberg, Pratyush Tandale, Roberto Fontanez, Oommen K Mathew, Andrew Kasarskis, Priyabrata Mukherjee, Lakshminarayanan Subramanian, Joel T Dudley, Khader Shameer, Sepsis in the era of data-driven medicine: personalizing risks, diagnoses, treatments and prognoses, Briefings in Bioinformatics, Volume 21, Issue 4, July 2020, Pages 1182–1195, https://doi.org/10.1093/bib/bbz059
[7] Arpaia, N.; Barton, G.M. The impact of Toll-like receptors on bacterial virulence strategies. Curr. Opin. Microbiol. 2013, 16, 17–22.
[8] Bulbul, I.J., Zahir, Z., Tanvir, A., Alam, Parisha, P. (2018). Comparative study of the antimicrobial, minimum inhibitory concentrations (MIC), cytotoxic and antioxidant activity of methanolic extract of different parts of Phyllanthus acidus (l.) Skeels (family: Euphorbiaceae). World Journal of Pharmacy and Pharmaceutical Sciences. 8(1):12-57. https://doi.org/10.20959/wjpps20191-10735
[9] Chaudhary, P.H.; Khadabadi, S. Bombax ceiba Linn.: Pharmacognosy, Ethnobotany and Phyto-pharmacology. Pharmacogn. Commun. 2012, 2, 2–9.
[10] Das K, Aminuzzaman FM (2017). Morphological and Ecological Characterization of Xylotrophic Fungi in Mangrove Forest Regions of Bangladesh. Journal of Advances in Biology & Biotechnology. 11(4):1–15.
[11] Das K, Aminuzzaman FM, Nasim A (2016). Diversity of Fleshy Macro Fungi in Mangrove Forest Regions of Bangladesh. Journal of Biology and Nature. 6(4):218–241.
[12] Dongmei Li, Xiaodong She, Richard Calderone, The antifungal pipeline: the need is established. Are there new compounds?, FEMS Yeast Research, Volume 20, Issue 4, June 2020, foaa023, https://doi.org/10.1093/femsyr/foaa023
[13] Jonathan M. Stokes, Kevin Yang, Kyle Swanson, et al. (2020). A Deep Learning Approach to Antibiotic Discovery. Cell. 180(4):688-702. https://doi.org/10.1016/j.cell.2020.01.021.
[14] Lee Y, Chou W, Chien T, Chou P, Yeh Y, Lee H. An App Developed for Detecting Nurse Burnouts Using the Convolutional Neural Networks in Microsoft Excel: Population-Based Questionnaire Study. JMIR Medical Informatics 2020;8(5):e16528
[15] Ma S, Chou W, Chien T, Chow J, Yeh Y, Chou P, Lee H. An App for Detecting Bullying of Nurses Using Convolutional Neural Networks and Web-Based Computerized Adaptive Testing: Development and Usability Study. JMIR mHealth and uHealth 2020;8(5):e16747
[16] Manik, M. M. T. G., Bhuiyan, M. M. R., Moniruzzaman, M., Islam, M. S., Hossain, S., & Hossain, S. (2018). The Future of Drug Discovery Utilizing Generative AI and Big Data Analytics for Accelerating Pharmaceutical Innovations, Nanotechnology Perceptions, Vol.14, No. 3 (2018), 120-135. https://doi.org/10.62441/nano-ntp.v14i3.4766
[17] Marzana A, Aminuzzaman FM, Chowdhury MSM, Mohsin SM, Das K (2018). Diversity and Ecology of Macrofungi in Rangamati of Chittagong Hill Tracts Under Tropical Evergreen and Semi-Evergreen Forest of Bangladesh. Advances in Research. 13(5):1–17.
[18] Meyer, V., Basenko, E.Y., Benz, J.P. et al. Growing a circular economy with fungal biotechnology: a white paper. Fungal Biol Biotechnol 7, 5 (2020). https://doi.org/10.1186/s40694-020-00095-z
[19] Miah, M. A., Rozario, E., Khair, F. B., Ahmed, M. K., Bhuiyan, M. M. R., & Manik, M. T. G. (2019). Harnessing Wearable Health Data and Deep Learning Algorithms for Real-Time Cardiovascular Disease Monitoring and Prevention, Nanotechnology Perceptions, Vol. 15 No. 3 (2019), 326-349. https://doi.org/10.62441/nano-ntp.v15i3.5278s
[20] Rosa A Börner, Vijayalakshmi Kandasamy, Amalie M Axelsen, Alex T Nielsen, Elleke F Bosma, Genome editing of lactic acid bacteria: opportunities for food, feed, pharma and biotech, FEMS Microbiology Letters, Volume 366, Issue 1, January 2019, fny291, https://doi.org/10.1093/femsle/fny291
[21] Rubina H, Aminuzzaman FM, Chowdhury MSM, Das K (2017). Morphological Characterization of Macro Fungi Associated with Forest Tree of National Botanical Garden, Dhaka. Journal of Advances in Biology & Biotechnology. 11(4):1–18.
[22] Sitarek, P.; Merecz-Sadowska, A.; Kowalczyk, T.; Wieczfinska, J.; Zajdel, R.; Śliwiński, T. Potential Synergistic Action of Bioactive Compounds from Plant Extracts against Skin Infecting Microorganisms. Int. J. Mol. Sci. 2020, 21, 5105. https://doi.org/10.3390/ijms21145105
[23] Tobore I, Li J, Yuhang L, Al-Handarish Y, Kandwal A, Nie Z, Wang L. Deep Learning Intervention for Health Care Challenges: Some Biomedical Domain Considerations. JMIR Mhealth Uhealth 2019;7(8):e11966