Received: 1 December 2023 | Revised: 31 January 2024 | Accepted: 18 February 2024 | Published online: 20 February 2024

REVIEW

Medinformatics
2024, Vol. 1(1) 2–9
DOI: 10.47852/bonviewMEDIN42022204

BON VIEW PUBLISHING

Artificial Intelligence with Great Potential in Medical Informatics: A Brief Review

Hao Lin[1,*]

[1]*School of Life Science and Technology, University of Electronic Science and Technology of China, China*

Abstract: In the 1950s and 1960s, in molecular biology, information technology was mainly applied to the molecular evolution of proteins and DNA and later expanded to multiple fields such as sequence alignment, protein structure prediction, and gene splicing. Entering the 21st century, the completion of the Human Genome Project marks the arrival of the era of biomedical big data, providing a large amount of data for the application of artificial intelligence in this field. Especially in recent years, the continuous accumulation of medical data has pushed the application of artificial intelligence in the medical field to a broader and more practical level. This paper briefly introduces the applications of artificial intelligence in genomics, proteomics, transcriptomics, epigenetics, drug development, and other fields. I hope this review can clearly introduce which biomedical fields artificial intelligence can be applied to and also promote doctors and related scholars to actively use artificial intelligence technology to solve specific biomedical problems.

Keywords: artificial intelligence, medical informatics, omics, prediction

1. Introduction

Since the lecture on "what is life" given by physicist Erwin Schrödinger in Dublin in 1943 (Schrödinger, 1944), biology has entered the molecular age. Afterward, in 1956, Watson and Crick unveiled the double helix structure of DNA, a breakthrough that marked the official commencement of molecular biology (Watson & Crick, 1953). The initiation of the Human Genome Project in 1990 (Wiechers et al., 2013) also pushed biomedical science into the information age.

Similarly, artificial intelligence (AI) can be traced back to Turing's vision of the "Turing machine" in the 1930s (Turing, 1936) and Fisher's linear discrimination during the same period (Fisher, 1936). Subsequently, at the workshop of the Dartmouth Summer Research Project on Artificial Intelligence in 1956, the assertion was made that "every aspect of learning or any other characteristic of intelligence should be accurately described so that machines can simulate it." This also marked the birth of AI.

In the early stages of development, these disciplines experienced numerous integrations, such as the application of AI in the structural identification of estrogen steroids (Smith et al., 1972) and the exploration of the relationship between drug mass spectrometry and pharmacological activity (Ting et al., 1973). In recent years, the emergence of various biotechnologies and the proliferation of scientific experiments have generated an enormous volume of data, opening up diverse applications for AI. The continuous advancement of deep learning technology and the ongoing enhancement of data processing capabilities have propelled the flourishing of AI in the field of biomedicine.

In this brief review, we will introduce some applications of artificial intelligence in the field of biomedicine. Based on different biomedical data sources, we will focus on the work of artificial intelligence in omics analysis.

2. Artificial Intelligence in Omics

2.1. Computational genomics

The genome, serving as a carrier of life information, has garnered widespread attention. Here, a brief overview will be presented based on research on the types of functional elements in the genome. Early bioinformatics analysis mainly focused on the identification of gene coding regions. Due to the fact that prokaryotic genes contain almost no introns, prediction for them is relatively easier for AI. There are some well-known prokaryotic coding region prediction tools, such as Glimmer (Salzberg et al., 1998) and ZCURVE (Guo et al., 2003). The features used in these algorithms are mainly based on the codon triplet encoding characteristics. Although most genes use ATG as the start codon, the mRNAs of many genes do not use the first ATG as the start codon. In addition, some genes use codon GTG as the start codon. Therefore, scholars designed AI-based models to identify translation initiation site (TIS) (Hirosawa et al., 1997). These models were constructed mainly using a purine-rich sequence upstream of the TIS. Good prediction performance for TIS was achieved in both prokaryotic and eukaryotic genomes (Sparks & Brendel, 2008). In addition, prokaryotes have the characteristic of transcribing multiple genes on one mRNA, so there are also some AI models for prokaryotic operons (transcription units) (Tomar et al., 2023).

In eukaryotes, the presence of introns leads to the emergence of broken genes, making gene prediction more difficult. Therefore, many scholars have designed various AI models to predict intron and exon

*Corresponding author: Hao Lin, School of Life Science and Technology, University of Electronic Science and Technology of China, China. Email: hlin@uestc.edu.cn

splice sites (Pertea et al., 2001). It is estimated that human has about 100,000 proteins; however, only about 15,000 genes were found in genome. This difference in quantity is caused by alternative splicing. Therefore, many AI models have emerged to predict alternative splicing event (Kan et al., 2001). However, due to the similarity of splicing site sequences and the difficulty in extracting nearby sequence information, the performance of sequence-based AI models is still far from satisfactory. Adding more regulatory information, such as epigenetic information and genomic structural information, may improve the predictive performance of AI models.

In the early stages of bioinformatics, there were relatively few AI models for gene transcription start sites (TSS), mainly due to the easy degradation of genes at the 5'UTR and 3'UTR ends after transcription into messenger RNA, resulting in insufficient data for training models. With the continuous accumulation of transcriptional data, research on TSS has also received attention. The transcription of genes is initiated by promoters; therefore, the prediction of TSSs and the prediction of promoters usually have the same significance. The differences in transcriptional regulation between eukaryotes and prokaryotes require the establishment of different predictive models. The prediction of eukaryotic promoters mainly targets the polII promoter in the human genome (Hannenhalli & Levy, 2001), which later extended to plants and other species (Bubnova et al., 2023). The prediction of promoters in prokaryotes mainly focused on the sigma70 promoter in Escherichia coli (Coppens & Lavigne, 2020) and later extended to the sigma54 promoter in prokaryotes (Liu et al., 2019). The predictive information of promoters mainly comes from sequence characteristics, mainly concentrated in several conserved regions that bind to RNA polymerase, such as TATA box, and purine-rich regions near TSS. The termination of gene transcription also plays an important role in gene expression; however, there are currently not many predictive models for transcription terminators (Feng et al., 2019). The reason why termination prediction is ignored may be because their functionality is not as important as transcription initiation.

The above is mainly studies on predicting different positions or functional regions in gene structure. There are also a large number of other functional elements in the genome, such as enhancers (Luo et al., 2023), silencers (Choy & Huang, 2002), and enzyme cleavage sites (Wang & Sun, 2023). Especially in recent years, due to the development of 3D genomes, enhancers have been found to play an important role in stabilizing the genome structure, which has attracted great attention (Beagrie et al., 2023). Many AI models for predicting enhancers based on sequence information have been established (Rapakoulia et al., 2023). In addition, research on enzyme cleavage sites on the genome mainly focuses on DNase hypersensitivity sites (Zhou et al., 2017), etc.

The genome is used to store genetic information, but the transmission of genetic information to the next generation is completed through replication. Most prokaryotes only require one origin of replication (ORI) to complete replication, making it relatively easy to recognize. The Z-Curve-based predictor has achieved good predictive performance for the ORI in prokaryotic genomes (Wang et al., 2021a). In addition, many models have been developed for the prediction of ORI in prokaryotes (Parikh et al., 2015). The genome of eukaryotes is relatively large, requiring multiple ORIs to complete genome replication, and has temporal and spatial specificity. Therefore, AI-based prediction of ORI in eukaryotes has also received widespread attention in recent years (Manavalan et al., 2021). These models not only obtain features through sequences but also combine epigenetic information and 3D genome information (Dao et al., 2022).

Of course, the information stored in the genome and the problems it brings go far beyond these. Some special genomic regions, such as Alu sequences (Torella et al., 2023), or some special types of genes, such as essential genes (Das & Sarkar, 2022), selenocysteine genes (Santesmasses et al., 2017), horizontal transfer genes (Sánchez-Soto et al., 2020), and recombination hotspot (Al Maruf & Shatabda, 2019), also have corresponding AI models. We will only provide a brief list here.

2.2. Computational proteomics

As a specific carrier of micro-level life activities, proteins have been extensively studied. AI models have also played a very important role in it. In fact, the study of protein informatics was slightly earlier than the study of genome informatics (Dayhoff, 1965).

Proteins are large molecules composed of amino acids that must be folded into specific three-dimensional structures in order to function. People divide the structure of proteins into multiple levels for research. The arrangement of amino acid residues in proteins is their primary structure. The secondary structure is a simple spatial structure formed by the shorter sequence fragments of a protein. Initially, the secondary structure of proteins can be classified into alpha helix, beta strand, and random coil and designed AI models to predict them (Peracha, 2024). Subsequently, some works classified the secondary structure of proteins into 8 categories and then predicted them using AI models (Li et al., 2023b). Some special secondary structures, such as beta-turn, attracted scholars' attention (Fang et al., 2020). The secondary structures of proteins are further assembled to form super-secondary structures, which have certain functions. Therefore, AI models were established to predict protein super-secondary structures (Anton et al., 2021). The actual functional entity is the tertiary structure of proteins. The prediction of the tertiary structure of proteins is the most concerned issue, and various prediction models have been developed to achieve the prediction of the tertiary structure (Zhang et al., 2010). In this field, a structure prediction competition called CASP has even been established, held once a year, to promote the theoretical analysis of protein structures (Das et al., 2023). However, for a considerable period of time, the accuracy of the model has been limited, and homologous modeling was the most reliable prediction method among them. Until the development of deep learning, AlphaFold developed by DeepMind under Google stood out in the 2018 CASP competition, significantly improving the accuracy of protein three-dimensional structure prediction and promoting the development of structural analysis, drug target research and development, and other fields (Lensink et al., 2023). Proteins often require the formation of polymers to function, so some studies have also built AI models to predict quaternary structure of proteins (Soltanikazemi et al., 2022).

In recent years, a phenomenon of phase separation in biological macromolecules was discovered, where some macromolecules gather together. The occurrence of these phase separation phenomena is closely related to intrinsic disordered proteins. Therefore, in recent years, the prediction of phase separation-related proteins has attracted scholars' attention (Lahorkar et al., 2023), and AI models for intrinsic disordered proteins have also received attention (Peng et al., 2020).

Proteins are not randomly distributed in cells, but need to reach specific positions to perform their unique functions. Therefore, a lot of work has been done for the prediction of subcellular localization of proteins, and various AI models have been designed (Özsarı et al., 2022). Due to species specificity, subcellular localization models are often designed based on species types, such as humans (Shen et al., 2020), plants (Sahu et al., 2020), other eukaryotes (Fink et al., 2006), Gram-negative bacteria (Romine, 2011), Gram-positive bacteria (Grasso et al., 2021), and viruses in host cells (Shen & Chou, 2007). Some AI models further focused on protein sub-organelle localization, such as subnuclear (Littmann et al., 2019),

submitochondria (Hou et al., 2021), and subchloroplast (Wang et al., 2023). Due to the fact that signal peptides guide proteins to specific cellular locations, AI models are also used for predicting protein signal peptides (Almagro Armenteros et al., 2019). The cell membrane plays an extremely important role in cells, responsible for material exchange and signal transmission inside and outside the membrane. Therefore, the recognition of membrane protein types has also received widespread attention (Sankari & Manimegalai, 2017).

The function of proteins is also reflected in their interactions; therefore, AI models of protein-protein interaction (PPI) have also been developed (Zhao et al., 2022). In addition, some special functional proteins have also been analyzed and modeled separately, such as hormone-binding protein (Butt et al., 2023) and cancerlectins (Tang et al., 2021). In order to search for drug targets, some AI models have also been developed for predicting G protein-coupled receptor (Nemoto et al., 2011) and ion channel (Gao et al., 2020).

Usually, those with longer residue sequences are called proteins, while those with shorter sequences are called peptides. Due to the convenience of peptides in design and application, they are more favored by scholars. As a potential therapeutic drug (Yan et al., 2022), toxin peptide (Monroe et al., 2023) has been developed with various AI models. In addition, there are corresponding AI models for other peptides, include antimicrobial peptide (Wang et al., 2016), anticancer peptide (Zhu et al., 2022), cell-penetrating peptide (Su et al., 2020), and blood-brain barrier penetrating peptide (Ma & Wolfinger, 2023).

Enzymes are the most widely used type of protein. Scholars have been striving to improve enzyme activity. AI models are also used for mutation of protein (Schomburg et al., 2017), thermal stability of enzyme (Li et al., 2022), enzyme suitability for acidic and alkaline environments (Zhang et al., 2009), solubility of enzyme (Wang et al., 2021c), and other aspects.

The above directions are basically based on the characteristics of amino acids to determine the type of proteins. In the field of diseases, by measuring the proteome of the disease group and the control group, AI models can be used to identify differential proteins and discover disease biomarkers (Chiam et al., 2015) and drug target (Dezső & Ceccarelli, 2020).

2.3. Computational transcriptomics

A transcriptome is a collection of all RNA transcribed by a specific tissue or cell during a certain developmental stage or functional state. AI models based on transcriptome data can provide assistance in various aspects of life development and disease diagnosis (Zhang et al., 2022). The basic idea behind it is that the gene expression levels vary among different tissues, especially for disease samples. The differentially expressed genes (DEG) or networks of DEG are considered one of the causes of disease occurrence and can serve as important biomarkers for disease diagnosis (Zolotareva et al., 2021).

Compared to the sequence information of genes or proteins, transcriptome data provide more noise, especially in sequencing data with severe batch effect (Luo et al., 2010). Directly using statistical methods to obtain DEGs is not reliable enough, and the prediction accuracy on independent samples is often unsatisfactory. Therefore, feature extraction methods based on gene expression order (Stretch et al., 2013) and gene expression network (Theofilatos et al., 2019) have also been proposed to discover true marker genes.

The gene expression obtained from bulk data is actually the average gene expression of each cell. Due to the emergence of single-cell technology, gene expression can be observed at the single-cell level (Goldman et al., 2019). In fact, most genes in a single cell are not expressed, so the gene expression matrix that appears is sparse. The most important aspect in studying single-cell transcriptomics is to determine the type of cell based on gene expression (Lin et al., 2017). Usually, scholars first perform cluster analysis on single-cell data, then determine the type of cells based on marker gene (Paisley & Liu, 2021), and then further analyze biological development and disease occurrence. Here, the marker gene is the foundation of single-cell analysis. In addition, it has been found that the results of bulk data analysis are often different from those obtained from single-cell data analysis. One of the current hot topics is how to use AI models to infer gene expression in single cells based on bulk data (Noureen et al., 2022).

2.4. Computational epigenetics

Scholars analogize DNA sequences to computer hardware, while epigenetic information is analogized to computer software. The DNA sequence in the genome is usually wrapped around nucleosomes, which are octamers formed by histones. Studies have found that nucleosomes prefer to appear in the exon region (Schwartz et al., 2009). The research of computational epigenetics initially focused on the construction of predictive models for nucleosome sequences (Teif, 2016). The model could predict which regions of the entire genome were more easily occupied by nucleosomes. Histones in nucleosomes undergo various chemical modifications. Therefore, histone modifications related to gene expression and disease occurrence are used to establish disease prediction models (Ho et al., 2012).

DNA, RNA, and proteins also undergo various chemical modifications, and the recognition of modification sites is of great significance for understanding the structure and function of macromolecules. Scholars have designed many AI models to identify modification sites in DNA, RNA, and proteins (Hasan et al., 2021). In addition, these modifications can also serve as characteristic information for biological development, species differentiation, and disease occurrence (Ao et al., 2021).

2.5. Other omics-related artificial intelligence research

AI has also been widely applied in various fields of omics. Next, we will provide a brief introduction.

Metabolomics is a discipline that conducts qualitative and quantitative analysis of all metabolites in a specific physiological period or biological process of an individual, organization, or cell (Tahir & Gerszten, 2020). Metabolomics has been applied to various stages of drug development, including drug target identification, drug metabolism analysis, drug response, and drug resistance research (Wishart, 2016). In addition, metabolomics has also been used to depict the evolutionary map of lung precancerous lesions to invasive lung adenocarcinoma (Nie et al., 2021). Using AI model, metabolomics biomarkers related to impaired fasting glucose and type 2 diabetes in large Chinese population can be obtained (Long et al., 2021).

Mass spectrometry is the most useful tool for metabolomics analysis, which can obtain a large number of metabolite ion characteristics from biological samples and generate rich metabolomics information (Krettler & Thallinger, 2021). Due to the vast variety and quantity of metabolites, mining and utilizing these mass spectrometry data is a challenge faced by AI. By conducting statistical analysis on metabolic data and using metabolic data as features to establish AI models, the relationship between metabolic molecules and physiological, pathological, and biological processes can be revealed (Wang et al., 2021b). In addition, it is also

necessary to establish AI models to annotate and identify metabolites (Yilmaz et al., 2020). Due to its high cost-effectiveness, metabolomics has been given high expectations by researchers.

The computational research on drugs encompasses many aspects, and here, we only focus on the works related to AI models. How to find drugs that can effectively bind to protein targets from a large number of compounds is the current direction of AI efforts (Liu et al., 2022). The prediction model for the new use of old drugs is also one of the hotspots in the field of drug research and development (Li et al., 2020). Drug-drug interaction (DDI) is crucial for the treatment of diseases. Various AI models have been developed to predict DDI (Nyamabo et al., 2022). Especially, the issue of drug combination therapy based on AI is of great concern to clinical practice (Yang et al., 2022). In addition, there are some inference models to predict the association between drugs and diseases (Guo et al., 2019), as well as the association between RNA and diseases (Yu et al., 2022).

Pathology is the "gold standard" for disease diagnosis. Computational pathology is a combination of digital pathology and AI technology (Campanella et al., 2019). Its emergence provides a guarantee for accurate and personalized treatment, diagnosis, and treatment. Computational pathology focuses on multiple data sources, such as pathological and tissue image information, using AI models to perform tasks such as detection, diagnosis, prediction, and prognosis, and showcasing clinically applicable knowledge to patients (Verghese et al., 2023). These studies on image processing, also known as imageomics (He et al., 2021), not only include pathological slice information but also X-ray images (Liu et al., 2022), magnetic resonance images (Bhalodiya et al., 2022), and so on.

Clinlabomics is a new concept proposed in recent years (Liu et al., 2018; Wen et al., 2022), typically characterized by large sample sizes, highly standardized data, but low feature dimensions. The information used in laboratory omics mainly comes from various tests conducted in medical biology, which can be demographic information such as age and gender, laboratory testing indicators such as liver function tests, urine tests, physiological indicators such as pulse, blood pressure, and even traditional Chinese medicine diagnostic indicators such as phlegm dampness and warm fever. The combination of these indicators with AI can provide assistance for disease warning and diagnosis and can also achieve disease occurrence analysis on a large population (Laursen et al., 2021). Therefore, there is a certain overlap between laboratory omics and the field of public health. For example, AI risk prediction model for diabetes and coronary heart disease was developed based on big data of physical examination (Li et al., 2023a; Meng et al., 2023; Yang et al., 2023).

3. Limitation

Presently, the application of AI in the field of biomedical research has yielded significant achievements. However, numerous studies continue to grapple with limitations imposed by sample constraints. Therefore, we will discuss some issues related to the samples.

Firstly, most AI-based models are built on prior knowledge. The effectiveness of these models mainly depends on the quality, quantity, and representativeness of the previous data. These prior data still heavily rely on specific wet experimental techniques.

Secondly, due to ethical limitations or technological barriers, many biomedical phenomena and mechanisms have not yet accumulated sufficient data, resulting in a lack of relevant artificial intelligence models.

Furthermore, there is a significant issue of data imbalance in the biomedical field. Although various sampling methods have been proposed to alleviate the problems of model bias and unreasonable evaluation caused by data imbalance, they still face challenges when the sample size is small.

Fourthly, the selection of control samples still troubles scholars with computational backgrounds. When selecting specific application scenarios and building models, careful selection of appropriate control samples is crucial for ensuring the practicality and reliability of subsequent models, which requires active participation from scholars with a deep understanding of computational principles in the biomedical context.

Finally, good data, especially long-term large queue data, requires year-round accumulation, not only careful planning by the initial designer but also the continuous efforts of multiple generations of researchers, which is often a difficult task to achieve.

4. Summary

The continuous updating of technology has generated various types of omics data, allowing us to observe life from various perspectives. Integrating these omics data and establishing AI models can better understand life phenomena and discover the laws of biological development. In fact, people have made a lot of attempts in this field. This review briefly explains the application of AI models in the biomedical field from omics perspective. In short, the fusion of multi-source information, interpretability, and high-performance AI models has achieved success in the biomedical field in recent years. However, we cannot be blinded by the current flourishing of the application of AI in biomedical data. We need to pay attention to potential defects and shortcomings in order to better develop this field.

Conflicts of Interest

Hao Lin is the Editor-in-Chief for *Medinformatics*, and was not involved in the editorial review or the decision to publish this article. The author declares that he has no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

References

Al Maruf, M. A., & Shatabda, S. (2019). iRSpot-SF: Prediction of recombination hotspots by incorporating sequence based features into Chou's Pseudo components. *Genomics*, *111*(4), 966–972. https://doi.org/10.1016/j.ygeno.2018.06.003

Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., …, & Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, *37*(4), 420–423. https://doi.org/10.1038/s41587-019-0036-z

Anton, B., Besalú, M., Fornes, O., Bonet, J., Molina, A., Molina-Fernandez, R., …, & Oliva, B. (2021). On the use of direct-coupling analysis with a reduced alphabet of amino acids combined with super-secondary structure motifs for protein fold prediction. *NAR Genomics and Bioinformatics*, *3*(2), lqab027. https://doi.org/10.1093/nargab/lqab027

Ao, C., Yu, L., & Zou, Q. (2021). Prediction of bio-sequence modifications and the associations with diseases. *Briefings in Functional Genomics*, *20*(1), 1–18. https://doi.org/10.1093/bfgp/elaa023

Beagrie, R. A., Thieme, C. J., Annunziatella, C., Baugher, C., Zhang, Y., Schueler, M., . . . , & Pombo, A. (2023). Multiplex-GAM: Genome-wide identification of chromatin contacts yields insights overlooked by Hi-C. *Nature Methods, 20*, 1037–1047. https://doi.org/10.1038/s41592-023-01903-1

Bhalodiya, J. M., Lim Choi Keung, S. N., & Arvanitis, T. N. (2022). Magnetic resonance image-based brain tumour segmentation methods: A systematic review. *Digital Health*, *8*, 20552076221074122. https://doi.org/10.1177/20552076221074122

Bubnova, A. N., Yakovleva, I. V., Korotkov, E. V., & Kamionskaya, A. M. (2023). In silico verification of predicted potential promoter sequences in the rice (*Oryza sativa*) genome. *Plants*, *12*(20), 3573. https://doi.org/10.3390/plants12203573

Butt, A. H., Alkhalifah, T., Alturise, F., & Khan, Y. D. (2023). Ensemble learning for hormone binding protein prediction: A promising approach for early diagnosis of thyroid hormone disorders in serum. *Diagnostics*, *13*(11), 1940. https://doi.org/10.3390/diagnostics13111940

Campanella, G., Hanna, M. G., Geneslaw, L., Miraflor, A., Werneck Krauss Silva, V., Busam, K. J., . . . , & Fuchs, T. J. (2019). Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature Medicine*, *25*(8), 1301–1309. https://doi.org/10.1038/s41591-019-0508-1

Chiam, J. T. W., Dobson, R. J. B., Kiddle, S. J., & Sattlecker, M. (2015). Are blood-based protein biomarkers for Alzheimer's disease also involved in other brain disorders? A systematic review. *Journal of Alzheimer's Disease*, *43*(1), 303–314. https://doi.org/10.3233/jad-140816

Choy, Y. S., & Huang, L. (2002). Experimental studies of a drumlike silencer. *The Journal of the Acoustical Society of America*, *112*(5), 2026–2035. https://doi.org/10.1121/1.1508779

Coppens, L., & Lavigne, R. (2020). SAPPHIRE: A neural network based classifier for σ70 promoter prediction in Pseudomonas. *BMC Bioinformatics*, *21*(1), 415. https://doi.org/10.1186/s12859-020-03730-z

Dao, F. Y., Lv, H., Fullwood, M. J., & Lin, H. (2022). Accurate identification of DNA replication origin by fusing epigenomics and chromatin interaction information. *Research, 2022*, 9780293. https://doi.org/10.34133/2022/9780293

Das, M. M., & Sarkar, K. (2022). Evaluation of machine learning classifiers for predicting essential genes in *Mycobacterium tuberculosis* strains. *Bioinformation*, *18*(12), 1126–1130. https://doi.org/10.6026/973206300181126

Das, R., Kretsch, R. C., Simpkin, A. J., Mulvaney, T., Pham, P., Rangan, R., . . . , & Westhof, E. (2023). Assessment of three-dimensional RNA structure prediction in CASP15. *Proteins: Structure, Function, and Bioinformatics*, *91*(12), 1747–1770. https://doi.org/10.1002/prot.26602

Dayhoff, M. O. (1965). Computer aids to protein sequence determination. *Journal of Theoretical Biology*, *8*(1), 97–112. https://doi.org/10.1016/0022-5193(65)90096-2

Dezső, Z., & Ceccarelli, M. (2020). Machine learning prediction of oncology drug targets based on protein and network properties. *BMC Bioinformatics*, *21*, 104. https://doi.org/10.1186/s12859-020-3442-9

Fang, C., Shang, Y., & Xu, D. (2020). A deep dense inception network for protein beta-turn prediction. *Proteins: Structure, Function, and Bioinformatics*, *88*(1), 143–151. https://doi.org/10.1002/prot.25780

Feng, C. Q., Zhang, Z. Y., Zhu, X. J., Lin, Y., Chen, W., Tang, H., & Lin, H. (2019). iTerm-PseKNC: A sequence-based tool for predicting bacterial transcriptional terminators. *Bioinformatics*, *35*(9), 1469–1477. https://doi.org/10.1093/bioinformatics/bty827

Fink, J. L., Aturaliya, R. N., Davis, M. J., Zhang, F., Hanson, K., Teasdale, M. S., . . . , & Teasdale, R. D. (2006). LOCATE: A mouse protein subcellular localization database. *Nucleic Acids Research*, *34*, D213–D217. https://doi.org/10.1093/nar/gkj069

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, *7*(2), 179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x

Gao, J., Wei, H., Cano, A., & Kurgan, L. (2020). PSIONplus^{m} server for accurate multi-label prediction of ion channels and their types. *Biomolecules*, *10*(6), 876. https://doi.org/10.3390/biom10060876

Goldman, S. L., MacKay, M., Afshinnekoo, E., Melnick, A. M., Wu, S., & Mason, C. E. (2019). The impact of heterogeneity on single-cell sequencing. *Frontiers in Genetics, 10*, 8. https://doi.org/10.3389/fgene.2019.00008

Grasso, S., van Rij, T., & van Dijl, J. M. (2021). GP^{4}: An integrated gram-positive protein prediction pipeline for subcellular localization mimicking bacterial sorting. *Briefings in Bioinformatics*, *22*(4), bbaa302. https://doi.org/10.1093/bib/bbaa302

Guo, F. B., Ou, H. Y., & Zhang, C. T. (2003). ZCURVE: A new system for recognizing protein-coding genes in bacterial and archaeal genomes. *Nucleic Acids Research*, *31*(6), 1780–1789. https://doi.org/10.1093/nar/gkg254

Guo, Z. H., Yi, H. C., & You, Z. H. (2019). Construction and comprehensive analysis of a molecular association network via lncRNA–miRNA–disease–drug–protein graph. *Cells*, *8*(8), 866. https://doi.org/10.3390/cells8080866

Hannenhalli, S., & Levy, S. (2001). Promoter prediction in the human genome. *Bioinformatics*, *17*, S90–S96. https://doi.org/10.1093/bioinformatics/17.suppl_1.s90

Hasan, M. M., Shoombuatong, W., Kurata, H., & Manavalan, B. (2021). Critical evaluation of web-based DNA N6-methyladenine site prediction tools. *Briefings in Functional Genomics*, *20*(4), 258–272. https://doi.org/10.1093/bfgp/elaa028

He, T., Fong, J. N., Moore, L. W., Ezeana, C. F., Victor, D., Divatia, M., . . . , & Wong, S. T. C. (2021). An imageomics and multi-network based deep learning model for risk assessment of liver transplantation for hepatocellular cancer. *Computerized Medical Imaging and Graphics*, *89*, 101894. https://doi.org/10.1016/j.compmedimag.2021.101894

Hirosawa, M., Sazuka, T., & Yada, T. (1997). Prediction of translation initiation sites on the genome of *Synechocystis* sp. strain PCC6803 by hidden Markov model. *DNA Research*, *4*(3), 179–184. https://doi.org/10.1093/dnares/4.3.179

Ho, S. M., Johnson, A., Tarapore, P., Janakiram, V., Zhang, X., & Leung, Y. K. (2012). Environmental epigenetics and its implication on disease risk and health outcomes. *ILAR Journal*, *53*(3–4), 289–305. https://doi.org/10.1093/ilar/ilx029

Hou, Z., Yang, Y., Li, H., Wong, K. C., & Li, X. (2021). iDeepSubMito: Identification of protein submitochondrial localization with deep learning. *Briefings in Bioinformatics*, *22*(6), bbab288. https://doi.org/10.1093/bib/bbab288

Kan, Z., Rouchka, E. C., Gish, W. R., & States, D. J. (2001). Gene structure prediction and alternative splicing analysis using genomically aligned ESTs. *Genome Research*, *11*(5), 889–900. https://genome.cshlp.org/content/11/5/889

Krettler, C. A., & Thallinger, G. G. (2021). A map of mass spectrometry-based in silico fragmentation prediction and

compound identification in metabolomics. *Briefings in Bioinformatics*, *22*(6), bbab073. https://doi.org/10.1093/bib/bbab073

Lahorkar, A., Bhosale, H., Sane, A., Ramakrishnan, V., & Jayaraman, V. K. (2023). Identification of phase separating proteins with distributed reduced alphabet representations of sequences. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *20*(1), 410–420. https://doi.org/10.1109/TCBB.2022.3149310

Laursen, S. B., Oakland, K., Laine, L., Bieber, V., Marmo, R., Redondo-Cerezo, E., . . . , & Stanley, A. J. (2021). ABC score: A new risk score that accurately predicts mortality in acute upper and lower gastrointestinal bleeding: An international multicentre study. *Gut*, *70*(4), 707–716. https://doi.org/10.1136/gutjnl-2019-320002

Lensink, M. F., Brysbaert, G., Raouraoua, N., Bates, P. A., Giulini, M., Honorato, R. V., . . . , & Wodak, S. J. (2023). Impact of AlphaFold on structure prediction of protein complexes: The CASP15-CAPRI experiment. *Proteins: Structure, Function, and Bioinformatics*, *91*(12), 1658–1683. https://doi.org/10.1002/prot.26609

Li, G., Buric, F., Zrimec, J., Viknander, S., Nielsen, J., Zelezniak, A., & Engqvist, M. K. (2022). Learning deep representations of enzyme thermal adaptation. *Protein Science*, *31*(12), e4480. https://doi.org/10.1002/pro.4480

Li, K., Du, Y., Li, L., & Wei, D. Q. (2020). Bioinformatics approaches for anti-cancer drug discovery. *Current Drug Targets*, *21*(1), 3–17. https://doi.org/10.2174/1389450120666190923162203

Li, L., Cheng, Y., Ji, W., Liu, M., Hu, Z., Yang, Y., . . . , & Zhou, Y. (2023a). Machine learning for predicting diabetes risk in western China adults. *Diabetology & Metabolic Syndrome*, *15*(1), 165. https://doi.org/10.1186/s13098-023-01112-y

Li, S., Yuan, L., Ma, Y., & Liu, Y. (2023b). WG-ICRN: Protein 8-state secondary structure prediction based on Wasserstein generative adversarial networks and residual networks with Inception modules. *Mathematical Biosciences and Engineering*, *20*(5), 7721–7737. https://doi.org/10.3934/mbe.2023333

Lin, P., Troup, M., & Ho, J. W. K. (2017). CIDR: Ultrafast and accurate clustering through imputation for single-cell RNA-seq data. *Genome Biology*, *18*(1), 59. https://doi.org/10.1186/s13059-017-1188-0

Liu, L.-R., Chiu, H.-W., Huang, M.-Y., Huang, S.-T., Tsai, M.-F., Chang, C.-Y., & Chang, K.-S. (2022). Using artificial intelligence to establish chest X-ray image recognition model to assist crucial diagnosis in elder patients with dyspnea. *Frontiers in Medicine*, *9*, 893208. https://doi.org/10.3389/fmed.2022.893208

Littmann, M., Goldberg, T., Seitz, S., Bodén, M., & Rost, B. (2019). Detailed prediction of protein sub-nuclear localization. *BMC Bioinformatics*, *20*(1), 205. https://doi.org/10.1186/s12859-019-2790-9

Liu, B., Han, L., Liu, X., Wu, J., & Ma, Q. (2019). Computational prediction of sigma-54 promoters in bacterial genomes by integrating motif finding and machine learning strategies. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *16*(4), 1211–1218. https://doi.org/10.1109/TCBB.2018.2816032

Liu, B., Papadopoulos, D., Malliaros, F. D., Tsoumakas, G., & Papadopoulos, A. N. (2022). Multiple similarity drug–target interaction prediction with random walks and matrix factorization. *Briefings in Bioinformatics*, *23*(5), bbac353. https://doi.org/10.1093/bib/bbac353

Liu, W., Zhang, Y., Huang, J., Tian, H., & Fu, W. (2018). The application value of testing omics in precision medicine (In Chinese). *International Journal of Laboratory Medicine*, *39*(9), 1025–1028.

Long, J., Yang, H., Yang, Z., Jia, Q., Liu, L., Kong, L., . . . , & Yan, D. (2021). Integrated biomarker profiling of the metabolome associated with impaired fasting glucose and type 2 diabetes mellitus in large-scale Chinese patients. *Clinical and Translational Medicine*, *11*(6), e432. https://doi.org/10.1002%2Fctm2.432

Luo, J., Schumacher, M., Scherer, A., Sanoudou, D., Megherbi, D., Davison, T., . . . , & Zhang, J. (2010). A comparison of batch effect removal methods for enhancement of prediction performance using MAQC-II microarray gene expression data. *The Pharmacogenomics Journal*, *10*(4), 278–291. https://doi.org/10.1038/tpj.2010.57

Luo, X., Li, Q., Tang, Y., Liu, Y., Zou, Q., Zheng, J., . . . , & Xu, L. (2023). Predicting active enhancers with DNA methylation and histone modification. *BMC Bioinformatics*, *24*(1), 414. https://doi.org/10.1186/s12859-023-05547-y

Ma, C., & Wolfinger, R. (2023). A prediction model for blood-brain barrier penetrating peptides based on masked peptide transformers with dynamic routing. *Briefings in Bioinformatics*, *24*(6), bbad399. https://doi.org/10.1093/bib/bbad399

Manavalan, B., Basith, S., Shin, T. H., & Lee, G. (2021). Computational prediction of species-specific yeast DNA replication origin via iterative feature representation. *Briefings in Bioinformatics*, *22*(4), bbaa304. https://doi.org/10.1093/bib/bbaa304

Meng, J., Huang, F., Shi, J., Zhang, C., Feng, L., Wang, S., . . . , & Wu, Y. (2023). Integrated biomarker profiling of the metabolome associated with type 2 diabetes mellitus among Tibetan in China. *Diabetology & Metabolic Syndrome*, *15*(1), 146. https://doi.org/10.1186/s13098-023-01124-8

Monroe, L. K., Truong, D. P., Miner, J. C., Adikari, S. H., Sasiene, Z. J., Fenimore, P. W., . . . , & Nguyen, H. B. (2023). Conotoxin prediction: New features to increase prediction accuracy. *Toxins*, *15*(11), 641. https://doi.org/10.3390/toxins15110641

Nemoto, W., Fukui, K., & Toh, H. (2011). GRIPDB-G protein coupled receptor interaction partners DataBase. *Journal of Receptors and Signal Transduction*, *31*(3), 199–205. https://doi.org/10.3109/10799893.2011.563312

Nie, M., Yao, K., Zhu, X., Chen, N., Xiao, N., Wang, Y., . . . , & Hu, Z. (2021). Evolutionary metabolic landscape from preneoplasia to invasive lung adenocarcinoma. *Nature Communications*, *12*(1), 6479. https://doi.org/10.1038/s41467-021-26685-y

Noureen, N., Ye, Z., Chen, Y., Wang, X., & Zheng, S. (2022). Signature-scoring methods developed for bulk samples are not adequate for cancer single-cell RNA sequencing data. *eLife*, *11*, e71994. https://doi.org/10.7554/eLife.71994

Nyamabo, A. K., Yu, H., Liu, Z., & Shi, J. Y. (2022). Drug–drug interaction prediction with learnable size-adaptive molecular substructures. *Briefings in Bioinformatics*, *23*(1), bbab441. https://doi.org/10.1093/bib/bbab441

Özsarı, G., Rifaioglu, A. S., Atakan, A., Doğan, T., Martin, M. J., Çetin Atalay, R., & Atalay, V. (2022). SLPred: A multi-view subcellular localization prediction tool for multi-location human proteins. *Bioinformatics*, *38*(17), 4226–4229. https://doi.org/10.1093/bioinformatics/btac458

Paisley, B. M., & Liu, Y. (2021). GeneMarkeR: A database and user interface for scRNA-seq marker genes. *Frontiers in Genetics*, *12*, 763431. https://doi.org/10.3389/fgene.2021.763431

Parikh, H., Singh, A., Krishnamachari, A., & Shah, K. (2015). Computational prediction of origin of replication in bacterial genomes using correlated entropy measure (CEM). *Biosystems*, *128*, 19–25. https://doi.org/10.1016/j.biosystems.2015.01.001

Peng, Z., Xing, Q., & Kurgan, L. (2020). APOD: Accurate sequence-based predictor of disordered flexible linkers. *Bioinformatics*, *36*, i754–i761. https://doi.org/10.1093/bioinformatics/btaa808

Peracha, O. (2024). PS4: A next-generation dataset for protein single-sequence secondary structure prediction. *BioTechniques*, *76*(2), 63–70. https://doi.org/10.2144/btn-2023-0024

Pertea, M., Lin, X., & Salzberg, S. L. (2001). GeneSplicer: A new computational method for splice site prediction. *Nucleic Acids Research*, *29*(5), 1185–1190. https://doi.org/10.1093/nar/29.5.1185

Rapakoulia, T., Lopez Ruiz de Vargas, S., Omgba, P. A., Laupert, V., Ulitsky, I., & Vingron, M. (2023). CENTRE: A gradient boosting algorithm for Cell-type-specific ENhancer-Target pREdiction. *Bioinformatics*, *39*(11), btad687. https://doi.org/10.1093/bioinformatics/btad687

Romine, M. F. (2011). Genome-wide protein localization prediction strategies for gram negative bacteria. *BMC Genomics*, *12*, S1. https://doi.org/10.1186/1471-2164-12-S1-S1

Sahu, S. S., Loaiza, C. D., & Kaundal, R. (2020). Plant-mSubP: A computational framework for the prediction of single-and multi-target protein subcellular localization using integrated machine-learning approaches. *AoB Plants*, *12*(3), plz068. https://doi.org/10.1093/aobpla/plz068

Salzberg, S. L., Delcher, A. L., Kasif, S., & White, O. (1998). Microbial gene identification using interpolated Markov models. *Nucleic Acids Research*, *26*(2), 544–548. https://doi.org/10.1093/nar/26.2.544

Sánchez-Soto, D., Agüero-Chapin, G., Armijos-Jaramillo, V., Perez-Castillo, Y., Tejera, E., Antunes, A., & Sánchez-Rodríguez, A. (2020). ShadowCaster: Compositional methods under the shadow of phylogenetic models to detect horizontal gene transfers in prokaryotes. *Genes*, *11*(7), 756. https://doi.org/10.3390/genes11070756

Sankari, E. S., & Manimegalai, D. (2017). Predicting membrane protein types using various decision tree classifiers based on various modes of general PseAAC for imbalanced datasets. *Journal of Theoretical Biology*, *435*, 208–217. https://doi.org/10.1016/j.jtbi.2017.09.018

Santesmasses, D., Mariotti, M., & Guigó, R. (2017). Computational identification of the selenocysteine tRNA (tRNA^{Sec}) in genomes. *PLoS Computational Biology*, *13*(2), e1005383. https://doi.org/10.1371/journal.pcbi.1005383

Schomburg, K. T., Nittinger, E., Meyder, A., Bietz, S., Schneider, N., Lange, G., . . . , & Rarey, M. (2017). Prediction of protein mutation effects based on dehydration and hydrogen bonding–A large-scale study. *Proteins: Structure, Function, and Bioinformatics*, *85*(8), 1550–1566. https://doi.org/10.1002/prot.25315

Schrödinger, E. (1944). *What is life? The physical aspect of the living cell and mind*. UK: Cambridge University Press.

Schwartz, S., Meshorer, E., & Ast, G. (2009). Chromatin organization marks exon-intron structure. *Nature Structural & Molecular Biology*, *16*(9), 990–995. https://doi.org/10.1038/nsmb.1659

Shen, H. B., & Chou, K. C. (2007). Virus-PLoc: A fusion classifier for predicting the subcellular localization of viral proteins within host and virus-infected cells. *Biopolymers*, *85*(3), 233–240. https://doi.org/10.1002/bip.20640

Shen, Y., Ding, Y., Tang, J., Zou, Q., & Guo, F. (2020). Critical evaluation of web-based prediction tools for human protein subcellular localization. *Briefings in Bioinformatics*, *21*(5), 1628–1640. https://doi.org/10.1093/bib/bbz106

Smith, D. H., Buchanan, B. G., Engelmore, R. S., Duffield, A. M., Yeo, A., Feigenbaum, E. A., . . . , & Djerassi, C. (1972). Applications of artificial intelligence for chemical inference. VIII. Approach to the computer interpretation of the high resolution mass spectra of complex molecules. Structure elucidation of estrogenic steroids. *Journal of the American Chemical Society*, *94*(17), 5962–5971. https://doi.org/10.1021/ja00772a005

Soltanikazemi, E., Quadir, F., Roy, R. S., Guo, Z., & Cheng, J. (2022). Distance-based reconstruction of protein quaternary structures from inter-chain contacts. *Proteins: Structure, Function, and Bioinformatics*, *90*(3), 720–731. https://doi.org/10.1002/prot.26269

Sparks, M. E., & Brendel, V. (2008). MetWAMer: Eukaryotic translation initiation site prediction. *BMC Bioinformatics*, *9*, 381. https://doi.org/10.1186/1471-2105-9-381

Stretch, C., Khan, S., Asgarian, N., Eisner, R., Vaisipour, S., Damaraju, S., . . . , & Baracos, V. E. (2013). Effects of sample size on differential gene expression, rank order and prediction accuracy of a gene signature. *PLoS ONE*, *8*(6), e65380. https://doi.org/10.1371/journal.pone.0065380

Su, R., Hu, J., Zou, Q., Manavalan, B., & Wei, L. (2020). Empirical comparison and analysis of web-based cell-penetrating peptide prediction tools. *Briefings in Bioinformatics*, *21*(2), 408–420. https://doi.org/10.1093/bib/bby124

Tahir, U. A., & Gerszten, R. E. (2020). Omics and cardiometabolic disease risk prediction. *Annual Review of Medicine*, *71*, 163–175. https://doi.org/10.1146/annurev-med-042418-010924

Tang, F., Zhang, L., Xu, L., Zou, Q., & Feng, H. (2021). The accurate prediction and characterization of cancerlectin by a combined machine learning and GO analysis. *Briefings in Bioinformatics*, *22*(6), bbab227. https://doi.org/10.1093/bib/bbab227

Teif, V. B. (2016). Nucleosome positioning: Resources and tools online. *Briefings in Bioinformatics*, *17*(5), 745–757. https://doi.org/10.1093/bib/bbv086

Theofilatos, K., Korfiati, A., Mavroudi, S., Cowperthwaite, M. C., & Shpak, M. (2019). Discovery of stroke-related blood biomarkers from gene expression network models. *BMC Medical Genomics*, *12*(1), 118. https://doi.org/10.1186/s12920-019-0566-8

Ting, K. L. H., Lee, R. C. T., Milne, G. W. A., Shapiro, M., & Guarino, A. M. (1973). Applications of artificial intelligence: Relationships between mass spectra and pharmacological activity of drugs. *Science*, *180*(4084), 417–420. https://doi.org/10.1126/science.180.4084.417

Tomar, T. S., Dasgupta, P., & Kanaujia, S. P. (2023). Operon Finder: A deep learning-based web server for accurate prediction of prokaryotic operons. *Journal of Molecular Biology*, *435*(14), 167921. https://doi.org/10.1016/j.jmb.2022.167921

Torella, A., Budillon, A., Zanobio, M., Del Vecchio Blanco, F., Picillo, E., Politano, L., . . . , & Piluso, G. (2023). Alu-mediated insertions in the DMD gene: A difficult puzzle to interpret clinically. *International Journal of Molecular Sciences*, *24*(11), 9241. https://doi.org/10.3390/ijms24119241

Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Journal of Mathematics*, 230–265. Retrieved from: https://www.wolframscience.com/prizes/tm23/images/Turing.pdf

Verghese, G., Lennerz, J. K., Ruta, D., Ng, W., Thavaraj, S., Siziopikou, K. P., ..., & Grigoriadis, A. (2023). Computational pathology in cancer diagnosis, prognosis, and prediction–present day and prospects. *The Journal of Pathology*, *260*(5), 551–563. https://doi.org/10.1002/path.6163

Wang, D., Lai, F. L., & Gao, F. (2021a). Ori-Finder 3: A web server for genome-wide prediction of replication origins in *Saccharomyces cerevisiae*. *Briefings in Bioinformatics*, *22*(3), bbaa182. https://doi.org/10.1093/bib/bbaa182

Wang, F., Liigand, J., Tian, S., Arndt, D., Greiner, R., & Wishart, D. S. (2021b). CFM-ID 4.0: More accurate ESI-MS/MS spectral prediction and compound identification. *Analytical Chemistry*, *93*(34), 11692–11700. https://doi.org/10.1021/acs.analchem.1c01465

Wang, G., Li, X., & Wang, Z. (2016). APD3: The antimicrobial peptide database as a tool for research and education. *Nucleic Acids Research*, *44*(D1), D1087–D1093. https://doi.org/10.1093/nar/gkv1278

Wang, L. S., & Sun, Z. L. (2023). iDHS-FFLG: Identifying DNase I hypersensitive sites by feature fusion and local-global feature extraction network. *Interdisciplinary Sciences: Computational Life Sciences*, *15*(2), 155–170. https://doi.org/10.1007/s12539-022-00538-8

Wang, X., Han, L., Wang, R., & Chen, H. (2023). DaDL-SChlo: Protein subchloroplast localization prediction based on generative adversarial networks and pre-trained protein language model. *Briefings in Bioinformatics*, *24*(3), bbad083. https://doi.org/10.1093/bib/bbad083

Wang, X., Liu, Y., Du, Z., Zhu, M., Kaushik, A. C., Jiang, X., & Wei, D. (2021c). Prediction of protein solubility based on sequence feature fusion and DDcCNN. *Interdisciplinary Sciences: Computational Life Sciences*, *13*(4), 703–716. https://doi.org/10.1007/s12539-021-00456-1

Watson, J. D., & Crick, F. H. C. (1953). Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, *171*(4356), 737–738. https://doi.org/10.1038/171737a0

Wen, X., Leng, P., Wang, J., Yang, G., Zu, R., Jia, X., ..., & Luo, H. (2022). Clinlabomics: Leveraging clinical laboratory data by data mining strategies. *BMC Bioinformatics*, *23*(1), 387. https://doi.org/10.1186/s12859-022-04926-1

Wiechers, I. R., Perin, N. C., & Cook-Deegan, R. (2013). The emergence of commercial genomics: Analysis of the rise of a biotechnology subsector during the human genome project, 1990 to 2004. *Genome Medicine*, *5*, 83. https://doi.org/10.1186/gm487

Wishart, D. S. (2016). Emerging applications of metabolomics in drug discovery and precision medicine. *Nature Reviews Drug Discovery*, *15*(7), 473–484. https://doi.org/10.1038/nrd.2016.32

Yan, K., Lv, H., Guo, Y., Chen, Y., Wu, H., & Liu, B. (2022). TPpred-ATMV: Therapeutic peptide prediction by adaptive multi-view tensor learning model. *Bioinformatics*, *38*(10), 2712–2718. https://doi.org/10.1093/bioinformatics/btac200

Yang, H., Luo, Y. M., Ma, C. Y., Zhang, T. Y., Zhou, T., Ren, X. L., ..., & Lin, H. (2023). A gender specific risk assessment of coronary heart disease based on physical examination data. *npj Digital Medicine*, *6*(1), 136. https://doi.org/10.1038/s41746-023-00887-8

Yang, Y., Gao, D., Xie, X., Qin, J., Li, J., Lin, H., ..., & Deng, K. (2022). DeepIDC: A prediction framework of injectable drug combination based on heterogeneous information and deep learning. *Clinical Pharmacokinetics*, *61*(12), 1749–1759. https://doi.org/10.1007/s40262-022-01180-9

Yilmaz, A., Ustun, I., Ugur, Z., Akyol, S., Hu, W. T., Fiandaca, M. S., ..., & Graham, S. F. (2020). A Community-based study identifying metabolic biomarkers of mild cognitive impairment and Alzheimer's disease using artificial intelligence and machine learning. *Journal of Alzheimer's Disease*, *78*(4), 1381–1392. https://doi.org/10.3233/JAD-200305

Yu, S., Wang, H., Liu, T., Liang, C., & Luo, J. (2022). A knowledge-driven network for fine-grained relationship detection between miRNA and disease. *Briefings in Bioinformatics*, *23*(3), bbac058. https://doi.org/10.1093/bib/bbac058

Zhang, G., Gao, J., & Fang, B. (2009). Discriminating acidic, neutral and alkaline enzymes with secondary structure amino acid composition (in Chinese). *Chinese Journal of Biotechnology*, *25*(10), 1508–1515.

Zhang, J., Wang, Q., Barz, B., He, Z., Kosztin, I., Shang, Y., & Xu, D. (2010). MUFOLD: A new solution for protein 3D structure prediction. *Proteins: Structure, Function, and Bioinformatics*, *78*(5), 1137–1152. https://doi.org/10.1002/prot.22634

Zhang, Y., Xia, R., Lv, M., Li, Z., Jin, L., Chen, X., ..., & Jin, S. (2022). Machine-learning algorithm-based prediction of diagnostic gene biomarkers related to immune infiltration in patients with chronic obstructive pulmonary disease. *Frontiers in Immunology*, *13*, 740513. https://doi.org/10.3389/fimmu.2022.740513

Zhao, N., Zhuo, M., Tian, K., & Gong, X. (2022). Protein–protein interaction and non-interaction predictions using gene sequence natural vector. *Communications Biology*, *5*(1), 652. https://doi.org/10.1038/s42003-022-03617-0

Zhou, W., Sherwood, B., Ji, Z., Xue, Y., Du, F., Bai, J., ..., & Ji, H. (2017). Genome-wide prediction of DNase I hypersensitivity using gene expression. *Nature Communications*, *8*(1), 1038. https://doi.org/10.1038/s41467-017-01188-x

Zhu, L., Ye, C., Hu, X., Yang, S., & Zhu, C. (2022). ACP-check: An anticancer peptide prediction model based on bidirectional long short-term memory and multi-features fusion strategy. *Computers in Biology and Medicine*, *148*, 105868. https://doi.org/10.1016/j.compbiomed.2022.105868

Zolotareva, O., Khakabimamaghani, S., Isaeva, O. I., Chervontseva, Z., Savchik, A., & Ester, M. (2021). Identification of differentially expressed gene modules in heterogeneous diseases. *Bioinformatics*, *37*(12), 1691–1698. https://doi.org/10.1093/bioinformatics/btaa1038

How to Cite: Lin, H. (2024). Artificial Intelligence with Great Potential in Medical Informatics: A Brief Review. *Medinformatics*. *1*(1), 2–9. https://doi.org/10.47852/bonviewMEDIN42022204