

Artificial Intelligence: Use in Clinical and Genomic Diagnostics

S Rehan Ahmad*

Assistant Professor, Dept. of Zoology,

Hiralal Mazumdar Memorial College for Women, Dakshineswar, Kolkata, West Bengal

*Corresponding Author E-mail: zoologist.rehan@gmail.com

Received: 10.01.2022 | Revised: 24.02.2022 | Accepted: 8.03.2022

ABSTRACT

The development of computer systems that are capable of carrying out tasks that typically require human intelligence is known as artificial intelligence (AI). Recent and quickly rising interest in medical AI applications is a result of AI software and technology improvements, especially deep learning algorithms and the graphics processing units (GPUs) that enable their training. While other AI subtypes have started to show similar promise in different diagnostic modalities, AI-based computer vision methods are poised to change image-based diagnostics in clinical diagnostics. Large and complicated genomic datasets are processed using a particular form of AI algorithm known as deep learning in various fields, such as clinical genomics. In this review, we first provide an overview of the primary categories of issues that AI systems are best adapted to address, followed by a description of the clinical diagnostic tasks that are aided by these solutions. Then, we concentrate on recently developed techniques for certain clinical genomics applications, such as variant calling, genome annotation and variant categorization, and phenotype-to-genotype correlation. We conclude by talking about the future potential of AI in individualized medicine applications, particularly for risk prediction in common complex diseases, as well as the issues, constraints, and biases that must be carefully addressed for the successful deployment of AI in medical applications, particularly those using data from genomics and human genetics.

Keywords: Machine learning, genomics, artificial intelligence, healthcare, gene technology.

INTRODUCTION

The interdisciplinary discipline of biology known as genomics focuses on the investigation of genome structure, function, mapping, and editing. The entire collection of an organism's DNA, including its genes, is known as its genome. We can divide genomics into numerous subgroups, including control,

structure, and function genomics. Machine learning and artificial intelligence have an impact on almost every business. Healthcare is not an exception. The business community has traditionally accepted innovations, and more and more researchers are focusing on developments in artificial intelligence.

Cite this article: Ahmad, S Rehan (2022). Artificial Intelligence: Use in Clinical and Genomic Diagnostics, *Emrg. Trnd. Nutr.* 1(2), 42-50. doi: <http://dx.doi.org/10.18782/2583-4606.111>

This article is published under the terms of the [Creative Commons Attribution License 4.0](https://creativecommons.org/licenses/by/4.0/).

One of these disciplines is genomics. Machine learning has a growing role to play in the development of this field. Researchers can analyze the growing volume of genetic picture data by combining deep learning and computer vision techniques. Machine learning models can handle computer vision tasks like semantic segmentation, image identification, and image extraction (Rahman et al., 2020). By combining machine learning and natural processing methods, a large volume of genomics-related material that can be discovered in publicly available scientific publications may be examined. Researchers may use this method to find solutions to issues like relationship extraction, information retrieval, or identifying named individuals.

Due to the massive amount of research being done in this area right now, some systems are suited for use with natural language processing tasks (Donepudi et al., 2020).

Background Information and Insights on AI and Genomics

The ability to read a person's genetic code, which controls how they behave, is made possible through DNA sequencing. History is provided by summarising the fundamental scientific principles along the pathway from DNA to RNA to protein. DNA comprises the basic matches In light of the four essential units (A, C, G, and T) known as nucleotides, A pairs with T and C pairs with G. People's chromosomes are divided into a total of 23 sets.

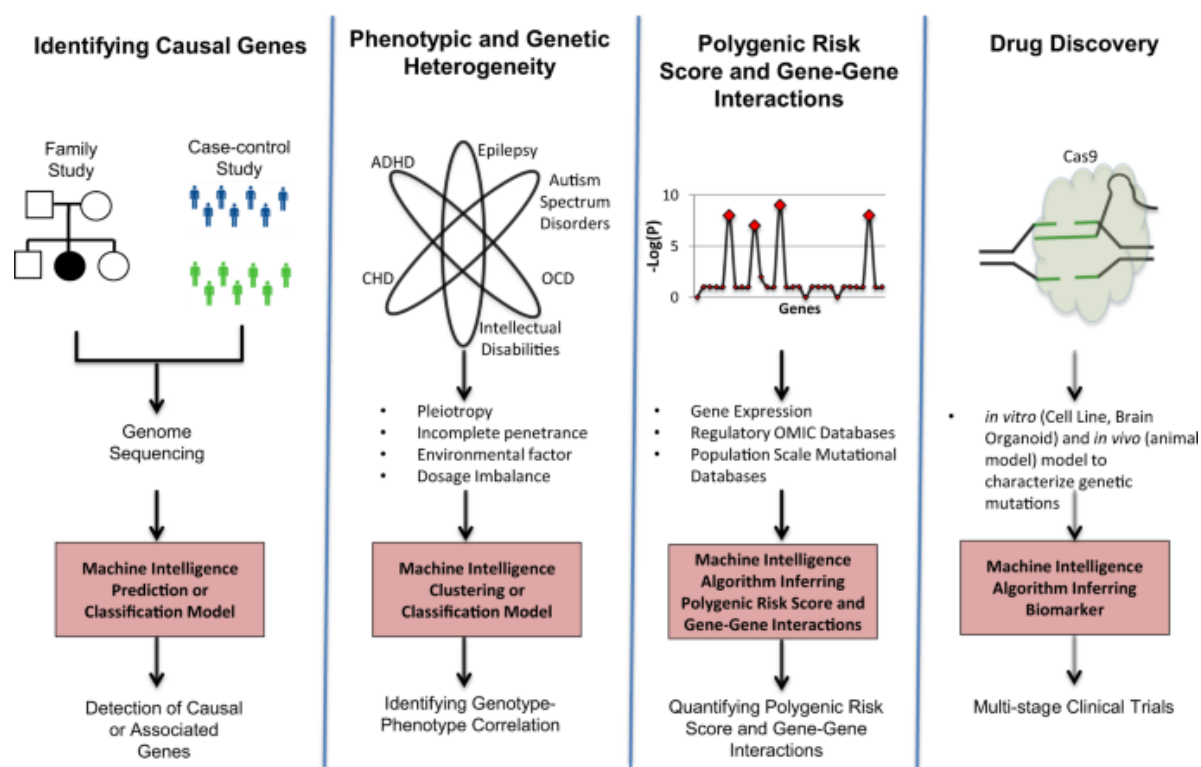


Figure 1: AI is the next big player in genomics

Chromosomes are more closely coordinated into DNA-based protein-producing or protein-encoding units known as genes. The number of attributes that a thing possesses is called its genome. A person has about 20,000 chromosomes and 3,000 billion base sets. Focusing on genomics research and business, just 2% of the human genome is used to encode proteins. It is a simple region. Genomic research is closely related to

Copyright © May-August, 2022; ETN

accuracy. Precision medicine, also known as tweaked medicine, is a field of medicine that integrates science, viewpoints, and networks with the aim of applying patient- or population-specific clinical mediation rather than a one-size-fits-all methodology. Its market size is predicted to reach \$87 billion by 2023. For instance, to reduce the likelihood of entanglements, a guy who requires blood bonding will be matched to a donor with a

similar blood categorization rather than a randomly chosen donor. High costs and infrastructural limitations are the two main barriers to the wider adoption of precision medicine. The enormous amount of patient data that needs to be collected and processed is being addressed by many researchers using machine learning techniques to cut costs. Fortunately, even with a significant relative fall in expenses between 2007 and 2012, the cost of decoding a genome tends to reduce annually for researchers and genomics companies.

AI and machine learning applications in genomics

The use of genetic testing and clinical administrations are affected by new machine

learning technologies in the field of genomics, which opens up the field to those interested in knowing how their ancestry may affect their well-being.

Gene sequencing

DNA sequencing is the process of determining the nucleic acid sequence or the order of the nucleotides in DNA. Any technique or technology that is utilized to determine the order of the four bases adenine, cytosine, guanine, and thymine is necessary. The field of whole genome sequencing (WGS) has become more popular in clinical diagnostics. Cutting edge sequencing has become a popular term that refers to cutting edge DNA sequencing techniques that aid in grouping by researchers.

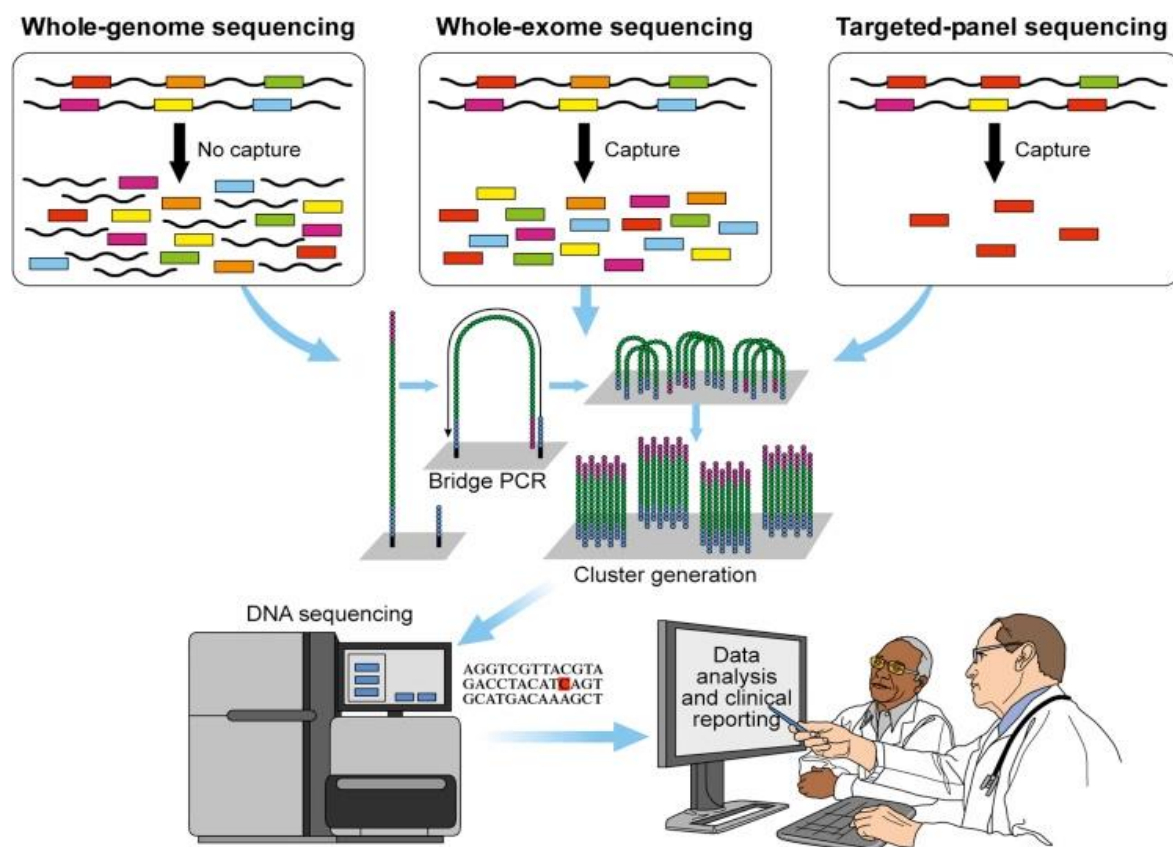


Figure 2: Genome sequencing

Organizations like Deep Genomics make use of machine learning to assist researchers in interpreting genetic variance. To help clients comprehend how crucial cellular processes are influenced by genetic diversity, algorithms are constructed based on patterns observed in large genetic data sets and then transformed into computer models. Cellular processes

Copyright © May-August, 2022; ETN

include, but are not limited to, DNA repair, metabolism, and cell formation. Theoretically, disruption of these pathways' normal function will result in disorders like cancer.

Three American venture capital firms have contributed \$3.7 million in seed funding to the 2014-founded Toronto-based enterprise. In fact, it's said that Deep Genomics' backers

advised the company to stay in Toronto and grow there rather than move to Silicon Valley. The choice will take into account the Canadian government's recent grant of \$125 million (Canadian dollars) for the Pan-Canadian Strategy for Artificial Intelligence. Deep Genomics has mentioned seven scientific publications as of April 2017, the majority of which predict or hint at potential genetic variations. However, this study's pertinent results have not yet been published in terms of diseases or potential treatments.

Editing the Gene

Gene editing is the process by which minute, exact alterations are achieved at the cellular level. The CRISPR system is thought to be the tool used for genome editing. It quickly and inexpensively completes editing. The researcher should select an appropriate aim sequence before using CRISPR. This complex system, which involves several choices and unpredictable outcomes, can be intimidating. The potential exists for machine learning to significantly reduce the time, price, and effort required to design a suitable sequence of goals.

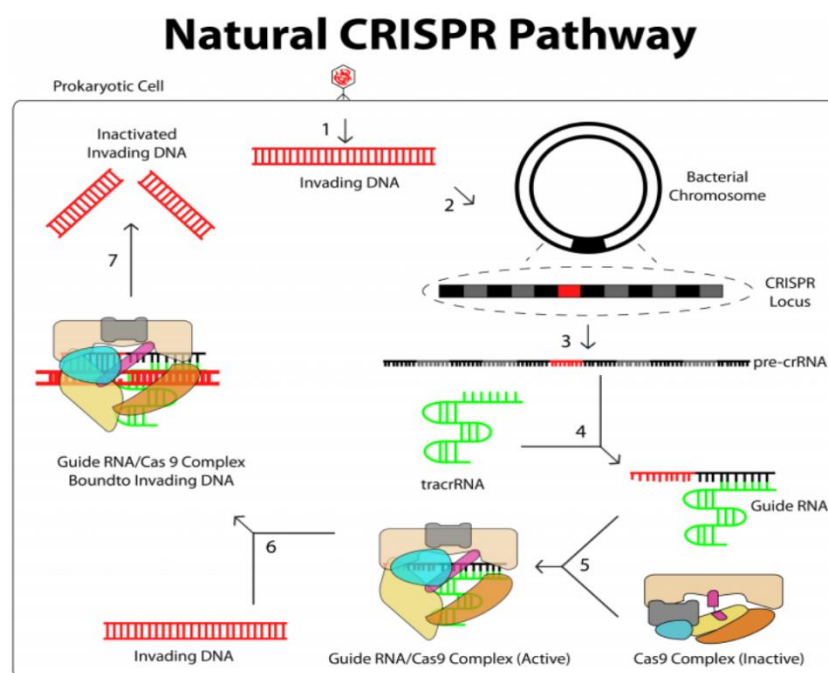


Figure 3: CRISPR gene editing

The London-based startup company Desktop Genetics is where AI and CRISPR meet. Since their formation in 2012, 7 investors have raised \$5.8 million in total equity investment, including accelerators, venture capital firms, and DNA sequencing pioneer Illumina. The company cites two important findings from a previous study, namely that the precision of an algorithm's capacity to predict CRISPR behaviour grows with an increase in the volume of training data and decreases when applied to a particular animal, such as mice versus humans. None of these findings was particularly surprising, and Desktop Genetics acknowledges that further investigation would be necessary to improve procedures and test

the limitations of how CRISPR can influence machine learning.

Pharmacogenomics

The technique focuses on how genes affect how an individual responds to medications. The relatively new field combines pharmacology (drug science) and genomics (genetic testing and its uses) to create trustworthy and safe medications and dosages that are tailored to a person's genetic makeup. Despite the fact that machine learning is still a very new science, there is evidence of studies in this area. For example, what is regarded as the first study to use machine learning algorithms to assess a safe dose of tacrolimus in patients with renal transplantation was

published in February 2017. Tacrolimus is typically administered to patients after successful organ transplantation in order to prevent acute rejection of new organs.

Methods for Newborn Screening

Experts anticipate that baby genetic screening will spread throughout society within the next ten years. The EHR of each person would be properly updated with information from the time of birth, and women who are breastfeeding would have access to non-intrusive screening tools for real issues like Down syndrome. The National Taiwan University Hospital's infant screening centre utilized AI to increase the accuracy of their online infant metabolic deformity screening system. Newborn genetic testing is becoming a more common practice. This non-invasive genetic screening can be used to identify diseases like Down syndrome at birth. Artificial intelligence can predict outcomes and the dangers involved in the treatment of genetic illnesses based on the evidence currently available.

For agriculture

The potential for genomics to improve soil quality and crop output in the agriculture sector is a new area of worry and optimism. Through its Illumina Accelerator programme, Illumina has provided financial support to business owners in California. To develop diagnostic tools for crop disease prediction and prevention, the startup is said to combine genomics and machine learning. The company is now called Trace Genomics and appears to have shifted its emphasis to soil health. If genetic data can be utilized to predict agricultural output or health, it may help farmers better forecast and maximize yields (and the resulting effects on soil). The global gains in crop yields may also increase the scale at which those improvements are applied, brought about by earlier genetic changes.

AI in Clinical Genomics

The AI algorithm aims to mimic human insight (Donepudi, 2017). However, when it comes to using traditional numerical techniques, AI in medical genomics tends to focus on tasks that are wasteful to do with

human expertise and defenceless against error. Numerous of the aforementioned methods, such as variant calling, genome explanation, variation marking, and correspondence from aggregate to genotype, have been altered to determine the various advancements connected with clinical genomic research. At some point, they may also be used to make predictions of genotype-to-aggregate. Here, we recognize the critical categorizations of problems investigated by AI in clinical genomics.

Variants Calling

Clinical genome understanding is helpless in the face of recognition, severe explicitness of individual hereditary variants within the huge numbers populating every genome, etc. Specific errors associated to test preparation nuances, sequencing innovation, grouping foundations, and the frequently unanticipated effects of science, such as physical mosaicism, are defenceless against standard variation-calling apparatuses (Li, 2014). Combining quantitative approaches with hand-created features, such as strand-predisposition or populace level conditions, was done to identify these problems, leading to high exactness but biased errors (DePristo et al., 2011).

AI computations may use a single genome with the perceived greatest quality level of reference variant calls to gather these inclinations and make unmatched variant calls. Deep Variant, a CNN-assembled variant caller that is legitimately prepared for perusing setups without any in-depth knowledge of genomics or sequencing stages, has recently been discovered to exceed the benchmark. (2018) Poplin et al. It is anticipated that the increased exactness results from CNNs' ability to recognize changing situations when sequencing information. Furthermore, ongoing studies demonstrate that deep learning can improve base-calling accuracy (and, consequently, variation distinguishing proof) for nanopore-based sequencing innovations that have traditionally attempted to compete with established sequencing innovation due to the error-prone nature of earlier base-calling algorithms (Wick et al., 2019).

Genome explanation and variant order

After variant calling, the analysis of the humanoid genome results is dependent on the distinct evidence provided by prior knowledge of certain hereditary variants and the suspicion of the influence of practical genomic elements on hereditary variation. Calculations based on artificial intelligence can encourage the use of earlier data by providing phenotype-to-genotype planning. Here, the identical amount of AI computations that were used to predict the presence of a useful component from crucial DNA grouping information is also employed to predict the effects of a hereditary minor deviation from such useful components, both genome explanation and variation.

Classification of coding variants

The nonsynonymous variants have been grouped in several ways (Tang & Thomas, 2016). When combined with either of these methods using relapse or other AI draws near, meta indicators (models that loop and aggregate the expectations generated by a few different indicators) that are focused on deep learning outperformed both their individual prescient segments and the combination of those prescient segments (Kircher et al., 2014). For instance, the combined annotation-based depletion approaches (CADD) in an AI system integrate different prophetic traits to forecast the harmfulness of inherited variants. A deep learning-based development of CADD known as DANN demonstrated improved performance using the same arrangement of informational components as CADD but combined in a profound neural organization (Quang et al., 2015). This specific development of CADD demonstrates how profound learning may be a very effective method for fusing well-known traits that indicate negative consequences.

Classification of noncoding variants

The algorithmic discovery and prediction of noncoding pathogenic variation is an unsolved problem in human genomics (Chatterjee & Ahituv, 2017). Recent research suggests that AI algorithms improve our capacity to interpret genetic variation that does not code. Splicing abnormalities in genes are caused by

at least 10% of uncommon harmful genetic mutations. However, they can be challenging to define due to the complexity of enhancers, silencers, isolators, and other combinatorial and long-range DNA interactions that affect the splicing of the genes (Soemedi et al., 2017).

Phenotype-to-genotype mapping

Independent of personal health state, a person's genetic composition contains a variety of inherited variants that have lately been identified as pathogenic or are anticipated to be harmful (Telenti et al., 2016). As a result, for a subatomic analysis of the disease, the identification of pathogenic fluctuations as well as the assurance of communication between the collective of the unwell creature and those predicted to happen from every incoming pathogenic variation, are also crucial. The mapping of phenotype to genotype may be greatly enhanced by computer-based intelligence algorithms, notably by removing more important level symptomatic rules implemented in clinical images and EHRs.

Genotype-to-phenotype prediction

The ultimate therapeutic goal of genetics is to incorporate forecasts and diagnoses of future illness risk. Risk classification for several frequent complex disorders is made possible by very simple statistical methods that are both clinically and personally valuable (Torkamani et al., 2018). Some studies have tried to depict unpredictable humanoid traits using AI algorithms gnomically. However, the majority of those documented so far in writing are likely to overfit because they allegedly depict significantly more variety in traits than should be practically based on heritability gauges. One application of AI to genetic stature expectation has the potential to provide sensible, reliable estimations within expected cutoff points, illustrative of the potential for AI-based to enhance computational approaches. However, combining multiple forms of health data and risk variables into potent infection hazard indicators is likely to be the benefit of AI-based approaches dealing

with genotype-to-phenotype expectation (Lello et al., 2018).

CONCLUSION

Machine learning in genomics is already changing a number of touchpoints, such as how genetic testing is conducted, how doctors provide medical care, and how easily consumers can access genomics to learn more about how their heredity can affect their health. With a greater interest in this subject, readers may want to study our recent piece on the uses of machine learning in medicine and pharma. Smart business is an attempt to implement AI to help accelerate the journey from bench to bedside and make precision medicine more widely used (Donepudi, 2018). Organizations that can provide real, workable solutions to precision medicine's difficulties may also benefit from such efforts. Despite the fact that there is a lot of hope, it is still tough to compete for precision medicine because many doctors are looking for additional information on the therapeutic value, and insurance companies are not seeing it as a need. Therefore, in addition to machine learning's capacity for data interpretation, education and clear examples of the value and significance of this technology would be required. Pharmacogenomics is a key area of the rapidly developing machine learning technology in genomics; however, this is just one example of many potential futures uses. With little empirical evidence, however, only time will show which industries will benefit most from investing in AI. We will keep a close eye on the field of genomics because it will be a busy one with more machine-learning applications in the near future.

Acknowledgements:

I take pride in acknowledging the insightful guidance of Principal, Dr. Soma Ghosh, Hiralal Mazumdar Memorial College for Women, Kolkata for giving me the inspiration to work on such innovative work. The author declared that the current paper was presented in the **National Seminar “Diversity in the Era of Globalization Challenges and Management”** organized by Bidhannagar
Copyright © May-August, 2022; ETN

College and sponsored by the Department of Higher Education, Government of West Bengal, on 22nd & 23rd February '2020”

ORCID id: <https://orcid.org/0000-0003-0796-5238>

Declarations:

Ethics approval and consent: This study has nothing to do with human and animal testing.

Consent for Publication: All the authors give their consent to publish the current manuscript.

Competing Interest: The authors declare that they have no conflict of interest.

Funding Declaration: The author declared that he hasn't received any rupee from any organization for this work.

REFERENCES

- Chatterjee S., & Ahituv, N. (2017). Gene Regulatory Elements, Major Drivers of Human Disease. *Annual Review of Genomics and Human Genetics*, 18, 45-63.
<https://doi.org/10.1146/annurev-genom-091416-035537>
- DePristo M. A., Banks E., Poplin R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernysky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43(5), 491–498.
<https://doi.org/10.1038/ng.806>
- Dolgin, E. (2019). AI face-scanning app spots signs of rare genetic disorders. *Nature*. <https://doi.org/10.1038/d41586-019-00027-x>
- Donepudi, P. K. (2017). Machine Learning and Artificial Intelligence in Banking. *Engineering International*, 5(2), 83-86.
<https://doi.org/10.18034/ei.v5i2.490>
- Donepudi, P. K. (2018). Application of Artificial Intelligence in Automation

- Industry. *Asian Journal of Applied Science and Engineering*, 7(1), 7–20. <http://doi.org/10.5281/zenodo.4146232>
- Donepudi, P. K., Ahmed, A. A. A., & Saha, S. (2020). Emerging Market Economy (EME) and Artificial Intelligence (AI): Consequences for the Future of Jobs. *PalArch's Journal of Archaeology of Egypt / Egyptology*, 17(6), 5562 - 5574. Retrieved from <http://palarch.nl/index.php/jae/article/view/1829>
- Gurovich, Y., Hanani, Y., & Bar, O. (2019). Identifying facial phenotypes of genetic disorders using deep learning. *Nature Medicine*, 25, 60-64. <https://doi.org/10.1038/s41591-018-0279-0>
- Hsieh, T. C., Mensah, M. A., & Pantel, J. T. (2019). PEDIA: prioritization of exome data by image analysis. *Genetics in Medicine*, 21, 2807–2814. <https://doi.org/10.1038/s41436-019-0566-2>
- Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*, 46(3), 310-315. <https://doi.org/10.1038/ng.2892>
- Köhler, S., Carmody, L., & Vasilevsky, N. (2019). Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Res.*, 47(D1), D1018-D1027. <https://doi.org/10.1093/nar/gky1105>
- Lello, L., Avery, S. G., Tellier, L., Vazquez, A. I., de Los Campos, G., & Hsu, S. D. H. (2018). Accurate Genomic Prediction of Human Height. *Genetics*, 210(2), 477-497. <https://doi.org/10.1534/genetics.118.301267>
- Li, H. (2014). Toward better understanding of artifacts in variant calling from high-coverage samples, *Bioinformatics*, 30(20), 2843–2851. <https://doi.org/10.1093/bioinformatics/btu356>
- Mobadersany, P., Yousefi S., Amgad, M., Gutman, D. A., Barnholtz-Sloan, J. S., Vega, J. E. V., Brat, D. J., & Cooper, L. A. D. (2018). Predicting cancer outcomes from histology and genomics using convolutional networks. *Proceedings of the National Academy of Sciences of the United States of America*, 115(13), E2970-E2979. <https://doi.org/10.1073/pnas.1717139115>
- Poplin, R., Chang, P.C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P. T., Gross, S. S., Dorfman, L., McLean, C. Y., DePristo, M. A. (2018). A universal SNP and small-indel variant caller using deep neural networks. *Nature Biotechnology*. 36(10), 983-987. <https://doi.org/10.1038/nbt.4235>
- Quang, D., Chen, Y., & Xie, X. (2015). DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics*, 31(5), 761-763. <https://doi.org/10.1093/bioinformatics/btu703>
- Rahman, M. M., Chowdhury, M. R. H. K., Islam, M. A., Tohfa, M. U., Kader, M. A. L., Ahmed, A. A. A., & Donepudi, P. K. (2020). Relationship between Socio-Demographic Characteristics and Job Satisfaction: Evidence from Private Bank Employees. *American Journal of Trade and Policy*, 7(2). 65-72. <https://doi.org/10.18034/ajtp.v7i2.492>
- Soemedi, R., Cygan, K. J., Rhine, C. L., Wang, J., Bulacan, C., Yang, J., Bayrak-Toydemir, P., McDonald, J., & Fairbrother, W.G. (2017). Pathogenic variants that alter protein code often disrupt splicing. *Nature Genetics*, 49(6), 848-855. <https://doi.org/10.1038/ng.3837>

- Tang, H., & Thomas, P. D. (2016). Tools for Predicting the Functional Impact of Nonsynonymous Genetic Variation. *Genetics*, 198(2), 685-691. <https://doi.org/10.1073/pnas.1613365113>
- Torkamani, A., & Wineinger, N. E. (2012). J. (2018). The personal and clinical utility of polygenic risk scores. *Nature Reviews, Genetics*, 19(9), 581-590. <https://doi.org/10.1038/s41576-018-0018-x>
- Telenti, A., Pierce, L. C., Biggs, W. H., di Iulio, J., Wong, E. H., Fabani, M. M., Kirkness, E. F., Moustafa, A., Shah, N., Xie, C., Brewerton, S. C., Bulsara, N., Garner, C., Metzker, G., Sandoval, E., Perkins, B. A., Och, F. J., Turpaz, Y., & Venter, J. C. (2016). Deep sequencing of 10,000 human genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 113(42), 11901-11906.
- Wick, R. R., Judd, L. M., & Holt, K. E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biology*, 20, 129. <https://doi.org/10.1186/s13059-019-1727-y>