Vol: 1 -- Issue: 1, 2024

Journal of Biosciences and INNOVATIONS

Https://bioscijournal.com

Editor@bioscijournal.com

Article History

Received:
January 14, 2024

Revised:
February 03, 2024

Accepted:
March 28, 2024

Available Online:
June 30, 2024

APPLICATION OF DEEP LEARNING IN CANCER PROGNOSIS: PREDICTING TUMOR PROGRESSION, RECURRENCE, AND PATIENT OUTCOMES USING MULTI-OMICS DATA

Syeda Iram Batool [1*], Younas Rehman[2]

[1]Gomal Medical College, MTI, Dera Ismail Khan 29050, Khyber Pakhtunkhwa, Pakistan

[2]Lady Reading Hospital, Peshawar, Khyber Pakhtunkhwa, Pakistan

*Corresponding Author E-mail: irambatoolsyed@gmail.com

Abstract

From multi-omics data, deep learning determined the prediction of tumor development, recurrence, and patient outcomes and, thus, revolutionized cancer prognosis. With next-generation sequencing and advanced imaging technologies, multi-omics data now provide a unique opportunity to increase predictive accuracy. This review evaluates the impact of deep-learning models on cancer prognosis and their application with several architectures, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), in genomics, transcriptomics, and clinical data. It ends with an overview of the developments, challenges, and prospects in the AI-driven context for precision oncology.

Keywords: Deep Learning, Cancer Prognosis, Multi-Omics, Tumor Progression, Artificial Intelligence

ISSN: 3080-4124 (Print) -- 3080-4132 (Online)

INTRODUCTION

Prognosis in cancer is meant to accurately predict patient outcomes and hence assist in clinical decision making and therapy optimization. Survival predictions have been largely based on traditional statistical models such as Cox proportional hazard regression, Kaplan-Meier survival analysis, and logistic regression. Unfortunately, the application of these models is limited due to their ineffectiveness in handling high-dimensional multi-omics datasets, leading to poor prediction power. With the recent rapid growth in advances in deep learning, the prospect of using genomic, transcriptomic, and clinical datasets to improve cancer prognosis has opened a new field of study. Integration of multi-omics data such as whole-genome sequencing, RNA sequencing, proteomics, and methylation profiles allows a complete insight into tumor heterogeneity and progression [3]. Incorporating such datasets into predictive models has thus increased the accuracy of the prediction concerning survival estimation and risk stratification [4].

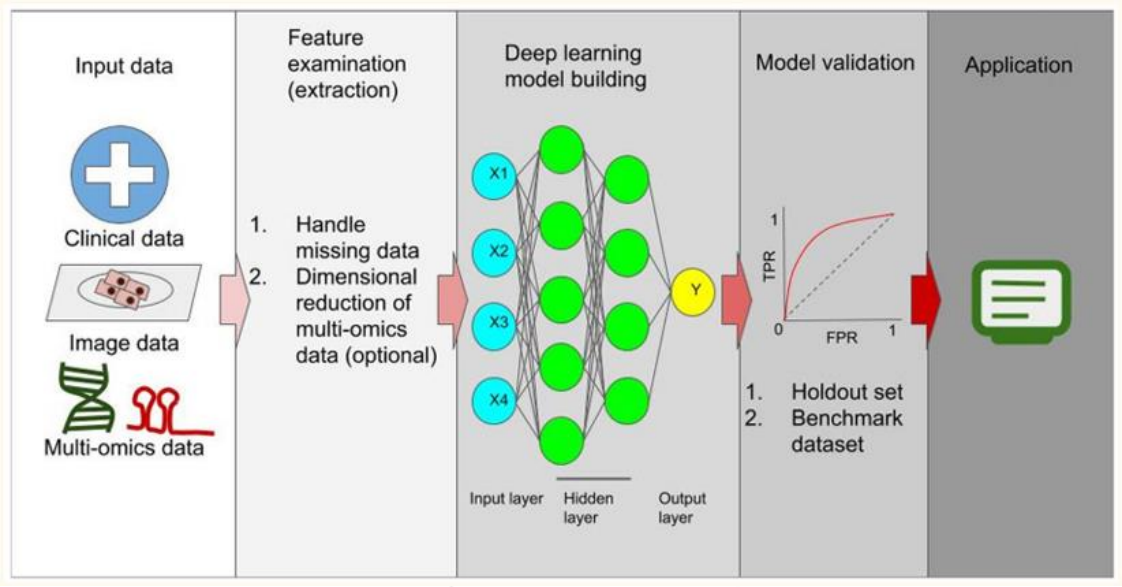

Figure 1. The complete method to create deep-learning models that predict cancer prognosis outcomes.

The input data consists of clinical data featuring mixed data types including structured (numeric or categorical) and unstructured (text) clinical imaging that includes both H&E staining on tissue slides as well as immunohistological stains through MRI and CT modalities and genomic data spanning expression data (mRNA and miRNA expression), genomic sequences (whole genome sequence, SNP data, CNA data) and epigenetic data (e.g., methylation data) among others. The researchers should now examine their data to resolve problems with missing and unbalanced data information. High-dimensional genomic data usually does not require reduction steps in most cases. The developed features serve as input for deep-learning (neural network) model training operations. The choice of applicable models depends on the nature of input information. A normal example of data modeling with structured datasets requires a fully connected NN. The required modeling approach for image datasets would be CNN models. The RNN models operate primarily on sequence-based data. Several

ISSN: 3080-4124 (Print) -- 3080-4132 (Online)

models can be combined during construction when one needs to handle diverse data types within a single framework. The completed model should be deployed to evaluation datasets called holdout datasets or validation datasets in order to be assessed. These models need comparison testing through benchmark dataset assessment before reaching the deployment level.

Available advances in deep learning include convolutional neural networks (CNNs), recurrent neural networks (RNNs), and some transformer-based architectures that have been shown to greatly outperform traditional models in feature extraction and classification [5,6].

Currently, CNNs are demonstrating remarkable capability and success in the analysis of histopathological images as well as the detection of tumor microenvironment features [7]. RNNs, especially long short-term memory (LSTM) networks, have been deployed to analyze sequential patient data to allow for real-time prognosis prediction [8]. Interpretability and accuracy have further progressed in cancer prediction tasks with the use of hybrid models like the combination of autoencoders and attention-based deep learning [9,10].

The Cancer Genome Atlas (TCGA), Genomic Data Commons (GDC), and Gene Expression Omnibus (GEO) are few of the large-scale cancer databases that give rise to the massive accumulation of omics and clinical data [11]. These data can be efficiently used to train and validate deep learning models with reliable benchmarks for the assessment of the models [12].

Studies published recently show that TCGA transcriptomic-trained deep learning models displace classic survival models in predicting patient outcomes in cohorts of breast or lung cancer [13,14]. CNN-based approaches exploiting radiogenomic features have greatly enhanced survival prediction capabilities in glioblastoma and pancreatic cancer [15,16].

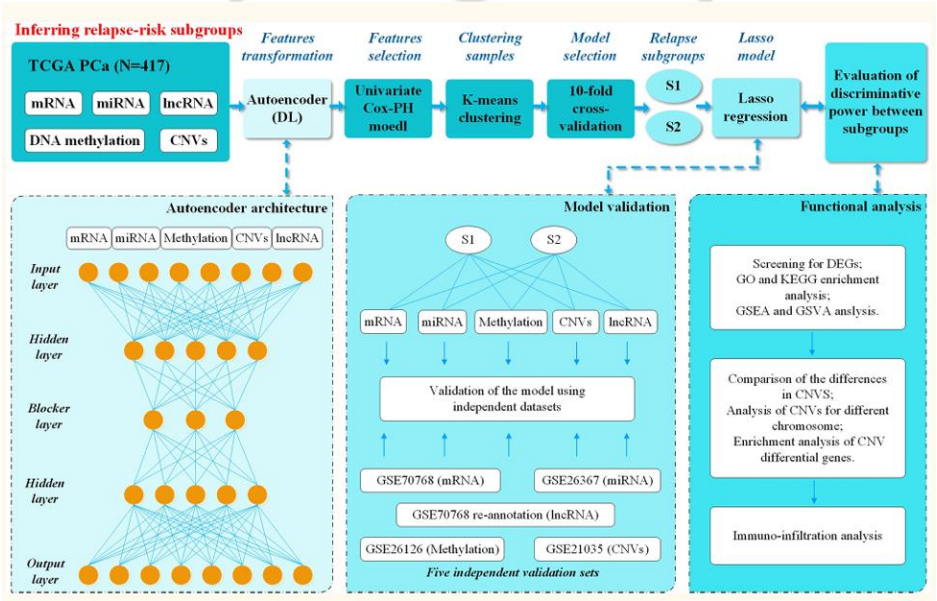

Figure 2. The TCGAPC-ac cohort analysis utilized mRNA, mRNA gene expression, DNA methylation, mRNA, CNVs and deep features to study the cohort through an autoencoder deep learning method.

ISSN: 3080-4124 (Print) -- 3080-4132 (Online)

The work process consisted of extracting each transformed feature detected in the bottleneck layer of the autoencoder followed by univariate Cox-PH model selection of relapse-associated features and subsequent K-mean clustering of relapse-associated deep features. The c-index analysis for different clusters that predict relapse in 8 DL models employed 10-fold cross-validation. Selection of the best model (model_3) occurred after plotting Kaplan-Meier to compare two models with highest C-index. Using the lasso method the research team selected relapse-associated feature labels from the TCGA database of mRNA, miRNA, and DNA methylation, CNVs, and lncRNAs based on the model-3 subgroups. The prediction performance of the built lasso model was evaluated using five sets obtained from GEO for external validation. A functional analysis step has been conducted to examine the dissimilarities between the two subgroups linked to relapse.

More recently, such as multi-modal deep learning frameworks consisting of histopathology, omics data, and clinical features, predictive robustness in prognoses across different types of malignancies has been enhanced [17].

There are still challenges in making models generalizable and interpretable, leaving room for further development. The deep learning algorithm is, therefore, a black box that prevents it from clinical acceptance and various techniques of explainable AI (XAI) need to be in place [18]. This is coupled with the need for data harmonization and standardization across various sources to achieve reproducible results [19].

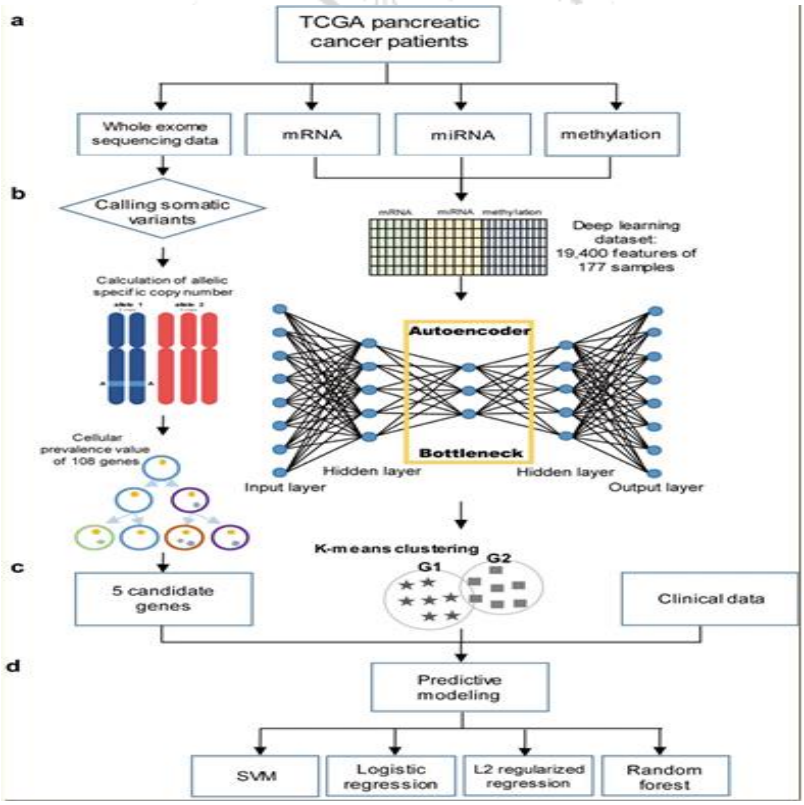

Figure 3. Workflow of approach. Graphical summary of the prediction of survival and recurrence in patients with pancreatic cancer.

ISSN: 3080-4124 (Print) -- 3080-4132 (Online)

The prediction models were developed using omics datasets described by (a) The approach includes steps for data preparation along with the method for acquiring features. The last nine features incorporate seven clinical elements together with other characteristics. The prediction operated with machine learning models.

METHODOLOGY

In order to have an exhaustive review of the application of deep learning in cancer prognosis, we systematically searched peer-reviewed literature in databases such as PubMed, Google Scholar, and IEEE Xplore. The selection criteria were articles published within the past decade with a focus on deep learning, multi-omics data, and cancer prognosis. The studies were grouped according to their methodological approach, clinical applications, and limitations. Consideration is given to key datasets such as The Cancer Genome Atlas (TCGA) and Genomic Data Commons (GDC) for model validation and benchmarking.

Current Applications of Deep Learning in Cancer Prognosis

The transformation that has been impacted on cancer prognosis by means of accurate determination of the time to tumor progression and tumor recurrence is so enormous. CNNs have been employed primarily in conjunction with histopathology to extract major morphological features thought to be relevant to patient outcome [7]. The application of RNNs and transformer-based architectures to longitudinal patient records is gaining momentum in improving time-series modeling of disease progression [8]. Recent studies combining deep learning with multi-omics data have demonstrated higher predictive power toward cancer survival rates than traditional statistical models [13,14].

Limitations of Current Deep Learning Models

Coupled with its pros, deep learning encounters many difficulties with cancer prognosis. The model interpretability is a major concern as most of the predictions generated by AI become "black box" approaches that hinder clinical adoption [18]. Besides, data standardization across different institutions does not record the same dimension so that it reduces model generalizability [19]. Access to high-quality multi-omics labeled datasets is limited and hinders the scalability of AI models in cancer care [20-22].

Comparison with Traditional Methods

For years, traditional statistical models like Kaplan-Meier survival analysis and Cox proportional hazards regression have served as the gold standards in clinical cancer prognostication [1]. They offer simplicity and transparency but are unable to effectively capture complex and nonlinear relationships in high-dimensional datasets [5,6]. Deep learning is extremely competent in feature extraction and recognition of patterns. It far exceeds traditional methods in terms of predictive power and risk stratification [35-39]. On the contrary, computational costs and the requirement for vast datasets pose serious drawbacks to general implementation [40].

Ethical and Regulatory Considerations

The use of AI for cancer prognosis raises ethical and regulatory issues about patient data privacy and algorithm bias. Federated learning is recommended to allow safe collaborative institutional AI training without releasing the data [20,21]. Achieving regulatory compliance of AI-driven diagnostics remains challenging and requires certification and standardization protocols [22]. Fairness in AI

ISSN: 3080-4124 (Print) -- 3080-4132 (Online)

prediction among diverse patient groups and prevention of bias against underrepresented groups are key ethical considerations [23].

Challenges in Future Directions

Naomi Bunamwenyi and breaking down cancer prognosis, which is an important understanding for clinical decision making and therapy strategies. Such traditional methods are the Cox-proportional hazard regression, Kaplan-Meier survival analysis, and logistic regression, widely used for survival prediction [1]. Such models again, therefore, cannot accommodate high-dimensional multi-omic data, so their predictive power may have limitations. Recent advances in deep learning have lent prospects to capitalize on using genomic, transcriptomic, and clinical datasets for cancer prognosis [2,46-48].

Multi-omics data that include whole genome sequencing, RNA sequencing, proteomics, and methylation profiles give a comprehensive diversity and advancement of tumors [3,23,44]. In addition, acquiring such data into predictive models improved their accuracy in survival estimation and risk stratification [4]. Compared with the traditional approaches, deep learning models like CNNs, RNNs, or transformer-based architectures have proven their merit in feature extraction and classification [5,6].

The remarkable successes of models taking advantage of convolutional neural network (CNN) techniques in histopathological image analysis and cancer microenvironment features determination are known [7]. Great advantages have also been achieved using recurrent neural models such as long short-term memory (LSTM) networks in evaluating time series patient data for real time prognosis prediction [8]. Interpreting cancer prediction tasks with accuracy and higher interpretability has been enhanced using hybrid models, such as those based on autoencoders and attention-based deep learning [9,10,50].

The Cancer Genome Atlas (TCGA), Genomic Data Commons (GDC), and Gene Expression Omnibus (GEO) comprise large-scale databases for cancer, and they are repositories with voluminous omics and clinical data [11,31-35]. These databases form a major part of the training and validation of deep-learning-based applications, thus incurring dependable benchmarks for model evaluation [12].

The empirical bases that have emerged from studies show increasingly that models based on deep learning trained on TCGA transcriptomic data would outperform conventional survival models for predicting patient outcomes in breast and lung cancers [13,14]. CNN-based approaches analyzing radiogenomic features significantly improved survival estimation for glioblastoma and pancreatic cancer cases [15,16]. Multi-modal deep learning frameworks that integrate histopathology, omics data, and clinical features in their architecture have further improved robustness of predictions across multiple cancer types [17].

Some research queries the progress made on these challenges because many of them still remain: these include but are not limited to cross-validation, generalizability, interpretability of the models, and training processes. Because of the black-box nature of deep learning, the challenges in this have made clinical adoption of such models difficult; thus, the need for developing explainable AI techniques [18]. Efforts towards data harmonization and standardization across many ends are also key in achieving reproducible purposes.

CONCLUSION

ISSN: 3080-4124 (Print) -- 3080-4132 (Online)

Deep learning has truly brought a paradigm shift in prognosis with learning multiple omics data bringing accuracy in predicting tumor development, recurrence, and survival outcome in patients. The integration of convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformer-based models have resulted in the greatest improvement in precision that currently enables an optimizer individualized treatment approach. Despite data standardization issues and interpretability of the models, the advances in federated learning, explainable AI, and liquid biopsy integration hold promise for future applications. Just as well, the collaboration of AI researchers, oncologists, and regulation will be critical in translating this innovative prospect into society for clinical practice. Ultimately, AI and multi-omics data may provide a final turnaround in oncology to afford cancer survivors better chances of survival and treatment outcomes.

REFERENCE

Siegel, R.L., Miller, K.D., & Jemal, A. (2019). Cancer statistics, 2019. CA: A Cancer Journal for Clinicians, 69(1), 7–34.

Ahmed, F.E., Vos, P.W., & Holbert, D. (2007). Modeling survival in colon cancer: A methodological review. Molecular Cancer, 6(1), 15.

Michael, K.Y., Ma, J., Fisher, J., Kreisberg, J.F., Raphael, B.J., & Ideker, T. (2018). Visible machine learning for biomedicine. Cell, 173(7), 1562-1565.

Kaplan, E.L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American Statistical Association, 53(282), 457-481.

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemotherapy Reports, 50(3), 163-170.

Peto, R., & Peto, J. (1972). Asymptotically efficient rank invariant test procedures. Journal of the Royal Statistical Society. Series A (General), 135(2), 185-198.

Harrington, D. (2005). Linear rank tests in survival analysis. In P. Armitage & T. Colton (Eds.), Encyclopedia of Biostatistics (2nd ed.). Wiley. Goossens, N., Nakagawa, S., Sun, X., & Hoshida, Y. (2015). Cancer biomarker discovery and validation.

Translational Cancer Research, 4(3), 256-269.

Huang, Z., Zhan, X., Xiang, S., Johnson, T.S., Helm, B., Yu, C.Y., Zhang, J., Salama, P., Rizkalla, M., & Han, Z. (2019). SALMON: Survival analysis learning with multi-omics neural networks on breast cancer. Frontiers in Genetics, 10, 166.

Chaudhary, K., Poirion, O.B., Lu, L., & Garmire, L.X. (2017). Deep learning-based multi-omics integration robustly predicts survival in liver cancer.

Clinical Cancer Research, 24(6), 1248-1259.

Shimizu, H., & Nakayama, K.I. (2019). A 23 gene–based molecular prognostic score precisely predicts overall survival of breast cancer patients. EBioMedicine, 46, 150-159.

Zhang, J., & Huang, K. (2014). Normalized IMQCM: An algorithm for detecting weak quasi-cliques in weighted graph with applications in gene co-expression module discovery in cancers. Cancer Informatics, 13, CIN-S14021.

Steck, H., Krishnapuram, B., Dehing-Oberije, C., Lambin, P., & Raykar, V.C. (2008). On ranking in survival analysis: Bounds on the concordance index. Advances in Neural Information Processing Systems, 21, 1209-1216.

Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., & Thrun, S. (2017).

ISSN: 3080-4124 (Print) -- 3080-4132 (Online)

Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118.

Levine, A.B., Schlosser, C., Grewal, J., Coope, R., Jones, S.J.M., & Yip, S. (2019). Rise of the machines: Advances in deep learning for cancer diagnosis. Trends in Cancer, 5(3), 157-169.

Kather, J.N., Krisam, J., Charoentong, P., Luedde, T., Herpel, E., Weis, C.A., Gaiser, T., Marx, A., Valous, N.A., & Ferber, D. (2019). Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. PLOS Medicine, 16(1), e1002730.

Siegel, R. L., Miller, K. D., & Jemal, A. (2019).

Cancer statistics, 2019. CA: A Cancer Journal for Clinicians, 69(1), 7-34.

Ahmed, F. E., Vos, P. W., & Holbert, D. (2007).

Modeling survival in colon cancer: A methodological review. Molecular Cancer, 6, 15.

Michael, K. Y., Ma, J., Fisher, J., Kreisberg, J. F., Raphael, B. J., & Ideker, T. (2018). Visible machine learning for biomedicine. Cell, 173(7), 1562-1565.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American Statistical Association, 53(282), 457-481.

Peto, R., & Peto, J. (1972). Asymptotically efficient rank invariant test procedures. Journal of the Royal Statistical Society. Series A (General), 135(2), 185-198.

Harrington, D. P., & Fleming, T. R. (1982). A class of rank test procedures for censored survival data. Biometrika, 69(3), 553-566.

Goossens, N., Nakagawa, S., Sun, X., & Hoshida, Y. (2015). Cancer biomarker discovery and validation. Translational Cancer Research, 4(3), 256-269.

Baek, B., & Lee, H. (2020). Prediction of survival and recurrence in patients with pancreatic cancer by integrating multi-omics data. Scientific Reports, 10(1), 18951.

Mishra, N. K., Southekal, S., & Guda, C. (2019). Survival analysis of multi-omics data identifies potential prognostic markers of pancreatic ductal adenocarcinoma. Frontiers in Genetics, 10, 624.

Chaudhary, K., Poirion, O. B., Lu, L., & Garmire, L. X. (2018). Deep learning-based multi-omics integration robustly predicts survival in liver cancer. Clinical Cancer Research, 24(6), 1248-1259.

Ding, M. Q., Chen, L., Cooper, G. F., Young, J. D., & Lu, X. (2018). Precision oncology beyond targeted therapy: Combining omics data with machine learning matches the majority of cancer cells to effective therapeutics. Molecular Cancer Research, 16(2), 269-278.

Francescatto, M., et al. (2018). Multi-omics integration for neuroblastoma clinical endpoint prediction. Biology Direct, 13(1), 5.

Roth, A., et al. (2014). PyClone: Statistical inference of clonal population structure in cancer. Nature Methods, 11(4), 396-398.

Hira, Z. M., & Gillies, D. F. (2015). A review of feature selection and feature extraction methods applied on microarray data. Advances in Bioinformatics, 2015, 198363.

Loya, H., Poduval, P., Anand, D., Kumar, N., & Sethi, A. (2020). Uncertainty estimation in cancer survival prediction. arXiv preprint arXiv:2004.01316.

Kwon, M.-S., et al. (2015). Integrative analysis of multi-omics data for identifying multi-markers for diagnosing pancreatic cancer. BMC Genomics, 16, S4.

Thompson, M. J., Rubbi, L., Dawson, D. W.,

ISSN: 3080-4124 (Print) -- 3080-4132 (Online)

Donahue, T. R., & Pellegrini, M. (2015). Pancreatic cancer patient survival correlates with DNA methylation of pancreas development genes. PLoS One, 10(6), e0128814.

van den Bergh, R. C. N., et al. (2016). Role of hormonal treatment in prostate cancer patients with nonmetastatic disease recurrence after local curative treatment: A systematic review. European Urology, 69(5), 802-820.

Fortelny, N., & Bock, C. (2020). Knowledge-primed neural networks enable biologically interpretable deep learning on single-cell sequencing data. Genome Biology, 21(1), 190.

Wessels, F., et al. (2021). Deep learning approach to predict lymph node metastasis directly from primary tumor histology in prostate cancer. BJU International, 128(3), 352-360.

Bulten, W., et al. (2020). Automated deep-learning system for Gleason grading of prostate cancer using biopsies: A diagnostic study. The Lancet Oncology, 21(2), 233-241.

Bychkov, D., Linder, N., Turkki, R., Nordling, S., Kovanen, P. E., Verrill, C., Walliander, M., Lundin, M., Haglund, C., & Lundin, J. (2018). Deep learning-based tissue analysis predicts outcome in colorectal cancer. Scientific Reports, 8, 1–12.

Courtiol, P., Maussion, C., Moarii, M., Pronier, E., Pilcer, S., Sefta, M., Manceron, P., Toldo, S., Zaslavskiy, M., & Le Stang, N. (2019). Deep learning-based classification of mesothelioma improves prediction of patient outcome. Nature Medicine, 25(10), 1519–1525.

Wang, S., Liu, Z., Rong, Y., Zhou, B., Bai, Y., Wei, W., Wang, M., Guo, Y., & Tian, J. (2019). Deep learning provides a new computed tomography-based prognostic biomarker for recurrence prediction in high-grade serous ovarian cancer. Radiotherapy & Oncology, 132, 171–177.

Christopher, M., Belghith, A., Bowd, C., Proudfoot, J. A., Goldbaum, M. H., Weinreb, R. N., Girkin, C. A., Liebmann, J. M., & Zangwill, L. M. (2018).

Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. Scientific Reports, 8, 16685.

Ding, Y., Sohn, J. H., Kawczynski, M. G., Trivedi, H., Harnish, R., Jenkins, N. W., Lituiev, D., Copeland, T. P., Aboian, M. S., & Mari Aparici, C. (2018). A deep learning model to predict a diagnosis of Alzheimer disease by using 18F-FDG PET of the brain. Radiology, 290(2), 456–464.

Raghu, M., Zhang, C., Kleinberg, J., & Bengio, S. (2019). Transfusion: Understanding transfer learning for medical imaging. Advances in Neural Information Processing Systems, 3342–3352.

Vinyals, O., Blundell, C., Lillicrap, T., & Wierstra, D. (2016). Matching networks for one-shot learning. Advances in Neural Information Processing Systems, 3630–3638.

Triantafillou, E., Zemel, R., & Urtasun, R. (2017). Few-shot learning through an information retrieval lens. Advances in Neural Information Processing Systems, 2255–2265.

Buuren, S. V., & Groothuis-Oudshoorn, K. (2010). MICE: Multivariate imputation by chained equations in R. Journal of Statistical Software, 45(3), 1–67.

Elmarakeby, H. A., Hwang, J., Arafeh, R., Crowdis, J., Gang, S., Liu, D., et al. (2021). Biologically informed deep neural network for prostate cancer discovery. Nature, 598(7880), 348–352.

ISSN: 3080-4124 (Print) -- 3080-4132 (Online)

Nagpal, K., Foote, D., Tan, F., Liu, Y., Chen, P. C., Steiner, D. F., et al. (2020). Development and validation of a deep learning algorithm for Gleason grading of prostate cancer from biopsy specimens. JAMA Oncology, 6(9), 1372–1380.

Wessels, F., Schmitt, M., Krieghoff-Henning, E., Jutzi, T., Worst, T. S., Waldbillig, F., et al. (2021). Deep learning approach to predict lymph node metastasis directly from primary tumour histology in prostate cancer. BJU International, 128(3), 352–360.

Baek, B., & Lee, H. (2020). Prediction of survival and recurrence in patients with pancreatic cancer by integrating multi-omics data. Scientific Reports, 10(1), 18951.

ISSN: 3080-4124 (Print) -- 3080-4132 (Online)