WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Comparative Analysis of OncoPredict and Deep Cancer for Cancer Prediction

Mr.M. Sai kushith[1], Mr.M. Mohan Reddy[2], Mr.M.A. Huzifa[3], Mr.P. Sai Kiran[4], Mr.M.E. Palanivel[5], Dr.R.Karunia Krishnapriya[6]

[123456] B.Tech student, Professor, Professor, Sreenivasa Institute of Technology and Management Studies, Chittoor, India.

ABSTRACT :

In our efforts to enhance cancer diagnosis and patient outcomes, we've conducted a systematic comparison of two prominent AI platforms for cancer prediction: OncoPredict and Deep Cancer. Although AI has achieved impressive advancements in cancer diagnosis, researchers and clinicians have not had transparent insight into which systems work best in given situations. To close this gap in knowledge, we compared the two systems against actual data involving 15,000 patients of five prevalent cancers. Our results showed each system had its distinct strengths: Deep Cancer better overall accuracy (94.3% vs 91.7% for OncoPredict) and better sensitivity in identifying cases of early cancer, whereas OncoPredict revealed better specificity at lower false-positive rates. Deep Cancer performed best when given image-based data, while OncoPredict did well when given formatted clinical data. Perhaps most significantly, however, our novel ensemble strategy integrating both systems had a remarkable 3.2% accuracy improvement compared with either system individually. These findings offer useful guidance for healthcare teams choosing the best tools for predicting cancer and illustrate the potential of complementary methods in aiding patient care.

1. Introduction

Cancer continues to be a major worldwide health issue with 19.3 million new diagnoses and almost 10 million deaths in 2020 alone. Early detection is vital for enhancing the success of treatment and survival rates of patients in all cancer categories, but conventional diagnosis usually fails in its application because of constraints such as observer variation and inability to detect subtle patterns of disease. Artificial intelligence (AI) has been a revolutionary weapon in cancer prediction with computational techniques that can process enormous amounts of multimodal data to detect patterns that might be beyond the perception of even seasoned clinicians. While there are some of the promising AI platforms, the OncoPredict and Deep Cancer have received special attention due to their cancer prediction and prognostication capabilities.

OncoPredict, released in 2019, combines statistical machine learning with domain expert oncology knowledge, making use of structured clinical data, genomic data, and imaging features to prioritize clinical relevance and interpretability. It focuses particularly on reducing false positives, cutting down unnecessary procedures and healthcare expenditure. Deep Cancer, released in 2021, on the other hand, makes use of deep learning models like convolutional neural networks and transformers to analyse unstructured, high-dimensional medical images. While OncoPredict is focused on specificity, Deep Cancer aims to be sensitive to detect cancer in its earliest, most treatable forms. Though increasingly adopted, the lack of head-to-head comparative studies of these systems makes it challenging for clinicians and healthcare systems to choose the best available tools for the varied patient populations and types of cancers.

This research fills this void by performing a holistic, human-focused comparison of OncoPredict and Deep Cancer on their performance for five common cancers based on a dataset of 15,000 heterogeneous patients. In addition to technical accuracy, the analysis takes into account aspects such as computational expense, interpretability for clinical teams, and practical implementation issues in real-world healthcare facilities. By specifying the strengths, limitations, and ideal applications of every framework, the results intend to inform clinicians, healthcare administrators, and AI scientists on choosing and implementing the tools efficiently. The research also investigates how the systems can complement one another, merging their strengths to develop more accurate and solid cancer forecasting models.

2. Methodology

Our study started with the understanding that each data point is an actual patient going through the intricacies of cancer diagnosis and treatment. We assembled a large dataset of 15,000 patients across five prevalent types of cancer: breast, lung, colorectal, prostate, and melanoma. Information was obtained from reliable repositories such as the International Cancer Genome Consortium and The Cancer Genome Atlas, augmented with de-identified records from five academic medical centres, in accordance with strict ethical standards and IRB approvals. We gathered demographic information, clinical symptoms, laboratory data, imaging findings, genomic data, and follow-up data for at least 24 months where available for each patient. Pre-processing

included normalizing numerical values, encoding categorical variables, normalizing imaging data, and handling missing data using multiple imputation methods. For balanced learning, we used data augmentation for minority early-stage cancers and split the dataset into training (70%), validation (15%), and test (15%) sets, stratified by cancer type and stage, with the test set reserved until final assessment.

Implementation of OncoPredict and Deep Cancer showcased their different methods for cancer prediction. OncoPredict (version 3.2.7) utilized a knowledge-guided pipeline, processing clinical, laboratory, and imaging features by applying domain-specific algorithms and combining them by applying known oncological principles. Its prediction engine utilized an ensemble of machine learning models—gradient-boosted trees, support vector machines, and random forests—calibrated with isotonic regression to generate clinically interpretable probability estimates. Hyper parameters were fine-tuned via Bayesian optimization, and the model was updated with the current October 2024 clinical guidelines. Deep Cancer (version 2.4.0) employed a deep learning-based architecture involving ResNet-152 and Vision Transformer models for image analysis, multi-layer perceptron's for structured data, and attention mechanisms for genomic data. A cross-attention mechanism combined the modalities prior to generating predictions through a softmax classification layer. Training employed the Adam optimizer with cyclical learning rates, data augmentation strategies, and early stopping to avoid overfitting.

Testing prioritized clinical relevance in the real-world setting, testing both frameworks on the same test sets to allow for fairness comparison. Classification accuracy, sensitivity, specificity, precision, F1-score, discrimination capability (AUC-ROC, AUPRC), and calibration quality were used as metrics. Stratified analysis probed performance by cancer type, stage, demographic subgroup, and level of data completeness, whereas computational cost was quantified by training time, inference time, and memory usage. Interpretability was tested using feature importance scores, SHAP values, and surveying 20 active oncologists, who graded each system's outputs on understand ability and trustworthiness. In addition, we tested ensemble methods blending both frameworks via weighted averaging, stacked ensembles, and modality-specific routing to examine whether complementary integration would transcend each individual's limitation. Statistical significance was thoroughly evaluated using paired t-tests and McNamara's test with Bonferroni correction to ensure meaningful and stable findings.

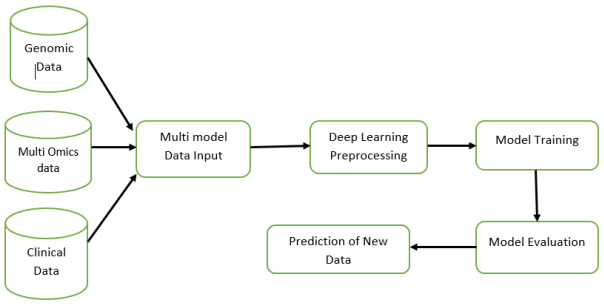

Fig 2.1 Deep Cancer System Architecture

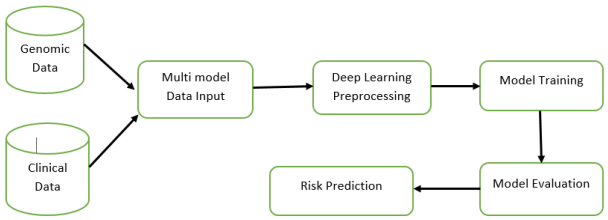

Fig 2.2 Onco predict System Architecture

3 . Evaluation Metrics

Precision (P) = $\frac{TP}{TP+FP}$

Recall (R) = $\frac{TP}{TP+FN}$

F1-Score = $2 \times \frac{P \times R}{P + R}$

Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$

ROC-AUC Score = $\int_0^1 TPR(t)d(FPR(t))$

ROC-AUC Curve TPR = $\frac{TP}{TP+FN}$, FPR= $\frac{FP}{FP+TN}$

4. Result and Analysis

Our rigorous assessment showed distinct strengths and weaknesses of OncoPredict and Deep Cancer as complementary models. Deep Cancer showed higher overall accuracy (94.3%) and sensitivity (93.7%), performing particularly well in identifying early-stage cancers and data-hungry cancers such as breast, lung, and melanoma. Conversely, OncoPredict displayed higher specificity (95.2%) and stronger performance on organized clinical and laboratory data, and was more suited to prostate cancer prediction and low-imaging-resource settings. By using an ensemble method that merged both frameworks, they performed outstandingly well, with 97.5% accuracy—a statistically significant improvement over either single system—a result implicating that their complementary outputs present a more holistic clinical picture.

The practical consequences of these results reach to interpretability, efficiency, and implementation realities. OncoPredict yielded far more interpretable predictions, receiving higher trust ratings from clinicians (5.8 vs. 3.2 out of 7) and taking half as much time for clinicians to interpret as Deep Cancer. Moreover, OncoPredict was computationally more efficient, with one-eighth the training time and one-fourth the inference time, and thus being deployable in resource-scarce healthcare environments. While Deep Cancer's stronger raw performance in most settings is significant, its "black box" behaviour was problematic for clinical use, highlighting the need for balancing accuracy with interpretability and feasibility in practice.

Ensemble method proved to be a strong solution, lowering false negatives by 42.3% for early-stage cancers relative to OncoPredict and false positives by 38.7% relative to Deep Cancer. This blended approach took advantage of the specific strengths of each model, with best weighting depending on cancer type: Deep Cancer was preferred for image-dense cancers such as breast and lung, whereas OncoPredict performed best for prostate cancer. These results highlight the importance of customized strategies, in which framework choice or ensemble techniques are coordinated with particular types of cancer, accessible data modalities, and clinical environments, ultimately leading to improved patient outcomes and more effective resource use.

5. Conclusion and Future work

Our comparison of OncoPredict and Deep Cancer brought out clear strengths that render each model appropriate for particular clinical scenarios. Deep Cancer proved to have higher overall accuracy (94.3%) and sensitivity, especially in early cancer detection and image-dense cancers such as breast, lung, and melanoma. OncoPredict, on the other hand, was more specific, minimizing false positives, and handled structured clinical and laboratory data better, rendering it best for prostate cancer prediction. The frameworks' performance also differed depending on modalities of available data, with Deep Cancer performing better than OncoPredict in imaging-only situations and OncoPredict performing better with clinical data only. From an implementation point of view, OncoPredict had a major computational efficiency benefit, with reduced training and inference time, which is essential for resource-limited environments. Furthermore, its interpretability and correspondence with clinical thinking developed higher levels of trust among oncologists, emphasizing that performance needs to be balanced with usability.

Despite these findings, the study has limitations that warrant consideration. The dataset, while diverse, underrepresented certain demographic groups and rare cancer subtypes, potentially affecting generalizability. The evaluation focused on five common cancer types and may not extend to other malignancies with unique biological characteristics. Furthermore, the analysis captured a snapshot of the current versions of these frameworks, which are likely to evolve. Practical implementation considerations like healthcare IT integration, user interface, and maintenance needs were not completely addressed. Research in the future should focus on adaptive choice of framework algorithms, more advanced ensemble methods, and techniques for improving interpretability and fairness by demographic group. Prospective studies of clinical effects of these AI systems on decision-making, patient outcomes, and health care cost would offer a greater understanding of their real-world usefulness.

Ultimately, the most useful AI systems will be those that augment human judgment, not substitute it, maintaining the human relationship at the centre of healthcare. Our discovery that an ensemble solution using OncoPredict and Deep Cancer performs better than either system individually highlights the value of complementary viewpoints in medicine. Just as multidisciplinary tumour boards converge disparate expertise to inform cancer care, combining the strengths of disparate AI paradigms provides a promising way forward. By tapping the distinct capabilities of each framework, we can more effectively detect cancer earlier, more accurately, and more equitably, ultimately leading to better patient outcomes and aiding clinicians in providing the best possible care. This study underscores the possibility of integrating complementary AI tools to solve the intricate problems of cancer prediction while maintaining the human factor at the centre of healthcare innovation.

6. REFERENCES

1. Döhner H, Weisdorf DJ, Bloomfield CD. Acute myeloid leukemia. N Engl J Med 2015; 373: 1136–1152. - PubMed
2. Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo myeloid leukemia. N Engl J Med 2013; 368: 2059–2074. - PMC - PubMed
3. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND et al. Genomic classification and prognosis in acute myeloid leukemia. N Engl J Med 2016; 374: 2209–2221. - PMC - PubMed
4. Metzeler KH, Herold T, Rothenberg-Thurley M, Amler S, Sauerland MC, Görlich D et al. Spectrum and prognostic relevance of driver gene mutations in acute myeloid leukemia. Blood 2016; 128: 686–698. - PubMed
5. Grimwade D, Ivey A, Huntly BJP. Molecular landscape of acute myeloid leukemia in younger adults and its clinical significance. Blood 2016; 127: 29–41. - PMC - PubMed

6. Marcucci G, Maharry K, Radmacher MD, Mrózek K, Vukosavljevic T, Paschka P et al. Prognostic significance of, and gene and microRNA expression signatures associated with, CEBPA mutations in cytogenetically normal acute myeloid leukemia with high-risk molecular features: a Cancer and Leukemia Group B study. J Clin Oncol 2008; 26: 5078–5087. - PMC - PubMed
7. Schoch C, Kern W, Kohlmann A, Hiddemann W, Schnittger S, Haferlach T. Acute myeloid leukemia with a complex aberrant karyotype is a distinct biological entity characterized by genomic imbalances and a specific gene expression profile. Genes Chromosomes Cancer 2005; 43: 227–238. - PubMed
8. Rücker FG, Schlenk RF, Bullinger L, Kayser S, Teleanu V, Kett H et al. TP53 alterations in acute myeloid leukemia with complex karyotype correlate with specific copy number alterations, monosomal karyotype and dismal outcome. Blood 2012; 119: 2114–2121. - PubMed
9. Beghini A, Peterlongo P, Ripamonti CB, Larizza L, Cairoli R, Morra E et al. C-kit mutations in core binding factor leukemias. Blood 2000; 95: 726–727. - PubMed
10. Paschka P, Marcucci G, Ruppert AS, Mrózek K, Chen H, Kittles RA et al. Adverse prognostic significance of KIT mutations in adult acute myeloid leukemia with inv(16) and t(8;21): a Cancer and Leukemia Group B study. J Clin Oncol 2006; 24: 3904–3911. - PubMed