ChatGPT in Predicting Progression-Free Survival in Prostate Cancer: Promise and Limitations

Engin Eren Kavak*

*Department of Medical Oncology, Etlik City Hospital, Oncology Hospital, Ankara, Turkey

INTRODUCTION

Recent advancements in Artificial Intelligence (AI) have precipitated a rapid transformation within various medical disciplines, most notably in the domains of predictive modelling and decision support systems. Within the field of oncology, for instance, there has been an escalating exploration of AI models that hold the potential to enhance survival predictions, guide treatment strategies, and improve patient care [1]. Among these models, ChatGPT, a large language model developed by OpenAI, has garnered considerable attention due to its capacity to process intricate medical data and provide estimated prognostic outcomes.

Despite the advancements mentioned above, the clinical reliability of AI-generated survival predictions remains a contentious issue. While traditional nomograms, such as Kattan and CAPRA-S scores, have long been utilised for prostate cancer prognosis, the potential of AI-driven tools in predicting Progression-Free Survival (PFS) is still in its infancy [2,3]. The recent study, titled "Progression-Free Survival Prediction Performance of ChatGPT: Analysis with Real-Life Data in Early and Locally Advanced Prostate Cancer," is one of the first attempts to systematically evaluate the performance of ChatGPT in predicting PFS in prostate cancer patients. This commentary aims to examine the findings critically, highlight their clinical implications, and discuss future directions for AI in oncologic prognosis.

ABOUT THE STUDY

The study by Kavak et al. concentrated on the evaluation of the accuracy of ChatGPT in predicting Progression-Free Survival (PFS) in 111 patients diagnosed with early-stage and locally advanced prostate cancer. The retrospective cohort included patients who had been followed for a minimum of 12 months. The survival estimates generated by ChatGPT were compared with the actual PFS times using statistical methods, including Kaplan-Meier survival analysis, Bland-Altman analysis, and paired t-tests.

KEY FINDINGS

ChatGPT Overestimates PFS: The AI model exhibited a tendency to overestimate PFS in comparison to the actual patient data. The mean discrepancy between ChatGPT's predictions and real-world PFS values was found to be 9.19 months [4]. The observed discrepancy was found to be statistically significant (p=0.048), however, its effect size (Cohen's d=0.189) was considered to be too small to have a significant impact on clinical decision-making [4].

Bias in AI Predictions: Bland-Altman analysis indicated a systematic bias (48.57 months), suggesting that ChatGPT's estimations were consistently skewed in one direction [4]. The predictive power of clinical variables is limited: traditional clinical risk factors (e.g., age, ECOG score, PSA levels, Gleason score) did not significantly influence recurrence prediction, implying that ChatGPT was unable to fully leverage these factors in its survival estimations [4].

Implications of the findings

The findings of this study underscore a pivotal challenge in the integration of AI-based survival models into the realm of clinical oncology. While ChatGPT has demonstrated an aptitude for the efficient processing and analysis of voluminous datasets, its predictive capabilities remain inferior to those of well-established prognostic models.

The potential strengths of AI-based prognostic tools are manifold.

Scalability: AI models have the capacity to process vast quantities of patient data in real time [5].

Personalised predictions: It is posited that, with further refinement, AI models have the potential to generate patient-specific survival estimates, tailored to individual risk factors [6].

Clinical decision support: AI-generated predictions have the potential to function as a secondary opinion for oncologists, thereby complementing existing clinical tools [7].

Correspondence to: Engin Eren Kavak, Department of Medical Oncology, Etlik City Hospital, Oncology Hospital, Ankara, Turkey, E-mail: engineren2000@yahoo.com

LIMITATIONS AND CHALLENGES

The study population was relatively small (n=111), which is a limitation as it restricts the generalisability of the results. It is well-documented that artificial intelligence models require larger and more diverse datasets for robust predictions [8].

Data Quality and Bias: AI models, including ChatGPT, have been observed to exhibit biases derived from the training data to which they are exposed. This has been shown to result in systematic over- or underestimation of survival times [9].

The clinical context is limited in scope. Current AI models are not equipped to account for treatment-related variables, post-treatment complications, or the nuances of real-world clinical decision-making [10].

The findings emphasise that, despite the potential of AI as an adjunct, it is not yet capable of replacing traditional clinical prognostic models. Further validation in multicentre, prospective studies is imperative for AI-based models to gain acceptance in the field of oncology.

CONCLUSION

This study provides significant insights into the role of AI, specifically ChatGPT, in survival prediction for prostate cancer. While the AI model demonstrated statistically significant discrepancies in predicting PFS, its clinical impact remains limited due to a systematic bias and insufficient precision.

To enhance the efficacy of AI in the field of cancer prognosis, future research endeavours must concentrate on the following aspects: The training of artificial intelligence models on a range of multi-institutional datasets to enhance the accuracy of predictions. The development of hybrid models that integrate artificial intelligence-based predictions with validated clinical nomograms.

The incorporation of supplementary risk factors, including but not limited to molecular biomarkers, genetic profiling, and treatment responses, is a methodology employed to refine predictions. It is imperative to ensure that AI models are transparent and interpretable, thus enabling oncologists to comprehend the methodology behind predictions.

Even though the function of ChatGPT in the domain of oncology decision support is still in a state of development, this study provides a significant body of data that will facilitate the refinement of AI-driven survival prediction tools. Through the implementation of additional enhancements, it is anticipated that AI will ultimately evolve into a standardised instrument within the field of precision oncology.

REFERENCES

1. Prabakaran DA, Ibrahim SA. Advancing Precision Oncology with Artificial Intelligence.

2. Greene KL, Meng MV, Elkin EP, Cooperberg MR, Pasta DJ, Kattan MW, et al. Validation of the Kattan preoperative nomogram for prostate cancer recurrence using a community based cohort: Results from cancer of the prostate strategic urological research endeavor (capsure). J Urol. 2004;171(6 Part 1): 2255-2259.

3. Cooperberg MR, Hilton JF, Carroll PR. The CAPRA-S score: A straightforward tool for improved prediction of outcomes after radical prostatectomy. Cancer. 2011;117(22):5039-5046.

4. Kavak EE, Dilli I. Progression-Free Survival Prediction Performance of ChatGPT: Analysis With Real Life Data in Early and Locally Advanced Prostate Cancer. Prostate. 2025:e24871.

5. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol. 2017;2(4).

6. Talukder N. Clinical decision support system: An explainable AI approach. 2024.

7. Zabaleta J, Aguinagalde B, Lopez I, Fernandez-Monge A, Lizarbe JA, Mainer M, et al. Utility of Artificial Intelligence for Decision Making in Thoracic Multidisciplinary Tumor Boards. J Clin Med. 2025;14(2):399.

8. Wilson A, Anwar MR. The Future of Adaptive Machine Learning Algorithms in High-Dimensional Data Processing. Int Trans Artif Intell. 2024;3(1):97-107.

9. Cross JL, Choma MA, Onofrey JA. Bias in medical AI: Implications for clinical decision-making. PLOS Digital Health. 2024;3(11):e0000651.

10. Moon I. Reliable and Trustworthy AI for Evidence-based Clinical Decision Support in Cancer Care (Doctoral dissertation, Massachusetts Institute of Technology). 2024.