EPiC Series in Computing

Volume 97, 2024, Pages 101–113

Proceedings of 36th International Conference on
Computer Applications in Industry and Engineering

Tensor decompositions in cancer study; A comprehensive review

Elia A. Movahed[1], Farnoosh Koleini[2], and Nasseh Tabrizi[3]

[1] University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, U.S.A
`eliamov@unc.edu`
[2] East Carolina University, Greenville, North Carolina, U.S.A.
`Koleinif20@students.ecu.edu`
[3] East Carolina University, Greenville, North Carolina, U.S.A.
`tabrizim@ecu.edu`

Abstract

Tensor decomposition techniques have gained significant attention in cancer research due to their ability to unravel complex and high-dimensional data structures. In this study, we comprehensively review the research trends from 2013 to 2023. Several themes are discussed, including the problems and challenges regarding cancer datasets, specifically image data and omics data. We also explore proposed tensor decomposition algorithms to tackle these challenges and their applications in different types of cancer, as well as the limitations and shortcomings of this field, which call for further research and development. Our objective is to investigate the application of tensor decomposition methods in cancer research. We first introduce the concept of tensors as multidimensional arrays and highlight their relevance in modeling cancer data. Subsequently, we discuss various tensor decomposition algorithms, such as Tucker decomposition and Canonical Polyadic decomposition, along with their advantages and limitations. This review aims to assist researchers interested in tensor decomposition techniques, which offer a valuable tool for analyzing complex and heterogeneous cancer data, enabling the discovery of hidden patterns and providing biological insights.

1 Introduction

Cancer, a pervasive and complex group of diseases, has emerged as one of the greatest challenges to global health in recent times. It affects people of all ages, races, and socioeconomic backgrounds, causing significant physical, emotional, and financial burdens on individuals and societies worldwide. The uncontrolled growth and spread of abnormal cells in the body, which define cancer, can manifest in various forms, including breast cancer, lung cancer, colon cancer, and many others [1]. Understanding cancer and finding effective treatments for it is of paramount importance due to several critical issues associated with the disease. Firstly, cancer is one of the leading causes of death globally, accounting for millions of deaths each year. It not only impacts the lives of individuals directly affected by the disease but also affects their families and communities.

K. Kambhampaty, G. Hu and I. Roy (eds.), CAINE 2023 (EPiC Series in Computing, vol. 97), pp. 101–113

Cancer datasets, including imaging and omics data, are valuable resources that provide insights into the characteristics of tumors, their behavior, and potential treatment options. In this article, we will be focusing on imaging data and omics data. The process of generating imaging data is commonly used to capture detailed images of tumors and surrounding tissues [2]. The following imaging techniques are used to generate imaging data: computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and single-photon emission computed tomography (SPECT). The challenges that arise regarding the analysis of imaging data include the large variability among the images and the complexity of extracting relevant information from them [3].

Omics data refers to large-scale datasets that provide a comprehensive view of various biological molecules and their interactions within a biological system. Omics data encompasses several fields, including genomics, transcriptomics, proteomics, and metabolomics, each focusing on a specific aspect of biology [4]. It also provides valuable insights into the molecular underpinnings of biological systems, including cancer. By studying and analyzing these large-scale datasets, researchers can gain a deeper understanding of the complexities of cancer biology. This knowledge can help in developing more targeted and personalized approaches for cancer diagnosis, treatment, and prevention [5] [4]. The challenges for omics data shows that integrating different variables and data types makes the data extremely complex [4]. Another challenge associated with omics data is that the clinical variables may be affected due to the complexity of what it resembles [6].

By analyzing omics data from cancer patients and comparing it to data from healthy individuals, researchers can identify specific genetic variations, gene expression signatures, protein modifications, or metabolic profiles that are indicative of cancer. These molecular alterations can serve as potential biomarkers for early detection, prognosis, and predicting treatment response in cancer patients. Through the analysis of omics data, researchers can uncover key molecular signatures that contribute to the development and progression of cancer, providing valuable insights for precision medicine and targeted therapeutic interventions [7]. In multi-view cancer studies that includes data, the stress of important information of tensor data without damaging the internal structure due to outlier damage is an important research topic. Correntropy would help because of its extent with dealing with tensor data. It is also a measure of the similarity in information theoretic learning (ITL) [8].

Tensor decomposition refers to mathematical techniques used to break down multidimensional data structures, known as tensors, into simpler components. A tensor can be thought of as a generalization of a matrix, extending its dimensions beyond two [9]. Common fields that tensor decomposition would be used are Canonical Polyadic (CP) decomposition or PARAFAC, and Tucker decomposition [9] [10]. CP decomposition represents a tensor as a sum of rank-1 tensors and it decomposes a tensor into a set of component matrices or vectors [11]. A tensor is divided into a core tensor and factor matrices along each mode using Tucker decomposition. This decomposition provides more flexibility in identifying intricate connections within the data.

Recently, tensor decompositions have found applications in various areas of cancer research. Specifically, they have been utilized to analyze gene expression data and identify distinct patterns or subtypes of cancer. Through decomposing the high-dimensional gene expression tensor, researchers can reveal hidden structures and relationships within the data. This deeper understanding of tumor heterogeneity enables the development of personalized treatment strategies, leading to improved outcomes for cancer patients [12]. Tensor decompositions have also been utilized in medical imaging to extract meaningful features from multi-modal imaging data, enabling improved tumor detection, classification, and tracking. Overall, tensor decomposition techniques offer valuable tools for analyzing complex cancer data sets, including imaging data and multi-omics data.

This paper provides a comprehensive review of research conducted from 2013 to 2023, focusing on the latest advancements in cancer studies utilizing tensor decompositions. The integration of imaging data with other multi-omics information has propelled the widespread adoption of tensor decomposition algorithms in cancer research. These algorithms have proven to be highly valuable in tackling the challenges posed by diverse data sources and have contributed to the identification of biomarkers, highlighting their crucial role in the field.

2 Comprehensive Review

The steps in this comprehensive review was to define the goals of the survey. These goals are described as follows:

- Identifying the problems and challenges regarding the cancer data sets such as image data and omics data.

- Identifying tensor decomposition algorithms employed to tackle these challenges and their applications in different cancer studies.

In this comprehensive review, we first searched and found the literature for publications using different search engines and databases of publications. The search query we used was (”tensor” AND ”decomposition”) AND (”cancer” AND ”study”) AND (”spatial” AND ”imaging”). This search query was used on many databases including IEEE Xplore, SpringerLink, ACM Digital Library. From these databases, we studied each publication and used that information to the main question that were mentioned in this comprehensive review. Figure 1 illustrates the number of publications utilizing tensor decomposition for cancer studies during the specified time frame.

The collection of papers examined in-depth the various challenges associated with cancer study datasets and proposed innovative solutions by leveraging different tensor decomposition methods. These papers delved into the complex nature of cancer data, shedding light on the hurdles researchers face when working with large-scale, high-dimensional datasets that encompass diverse variables and intricate relationships.

One significant challenge that these papers addressed was the issue of data sparsity. To tackle this challenge, researchers explored tensor decomposition techniques that effectively handle missing data, enabling them to impute or predict the missing values accurately. By leveraging advanced tensor decomposition methods, such as Tucker decomposition or CANDECOMP/PARAFAC (CP) decomposition, the papers offered promising approaches to enhance data completeness and reliability [13].

In this comprehensive review paper, we delve into recent and innovative approaches that make significant contributions to the advancement of cancer research. These approaches not only aid in diagnosis and treatment but also have the potential to greatly improve patient outcomes. By exploring cutting-edge methods and techniques, this paper sheds light on promising avenues for tackling the challenges in cancer study datasets and highlights their potential impact on the field of oncology.

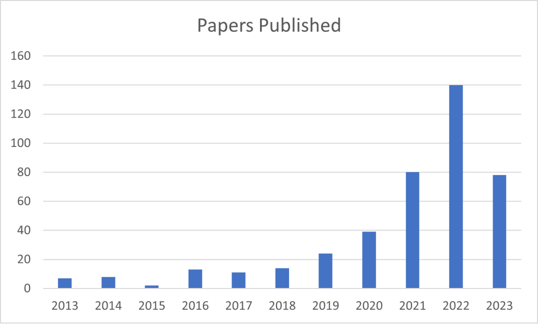


Figure 1: Number of papers published from 2013 to 2023

3 Cancer Study Data Types

Cancer research relies on various types of data to gain insights into the disease and develop better diagnostic and treatment approaches. Two important types of data used in cancer studies are image data and omics data.

3.1 Image Data

Image data in cancer studies refers to visual representations of tissues, cells, or organs obtained through medical imaging techniques such as X-rays, computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and histopathology slides. This type of data plays a critical role in cancer research as it allows researchers to observe and analyze the structural and functional characteristics of tumors and surrounding tissues [1]. Image data, such as radiology images or digital pathology images, plays a crucial role in cancer diagnosis and treatment. Clinicians rely on these images to detect and diagnose cancer, as they provide valuable information about the presence, location, size, and characteristics of tumors [12]. Under imaging, the topic of multilayer decomposition is important because it leverages a specific imaging layer in addition to a specific tensor structure. This method offers the advantage of simultaneously combining multimodality information and efficiently capturing the varied spatial aspects of signals without relying on a population structure [3]. Cancer images often exhibit high variability in terms of image quality, acquisition parameters, and imaging modalities. This variability poses challenges in standardizing the images and ensuring consistency across different datasets. To overcome these challenges and enhance the reliability of the analysis, preprocessing and normalization techniques are crucial [6] [14].

Tensor decompositions can help address several challenges in cancer image analysis. Here are some specific challenges that tensor decompositions can help overcome:

Heterogeneity: Tensor decompositions enable the extraction of latent features from multi-dimensional cancer image data, facilitating the identification of underlying patterns and subtypes. By decomposing the tensor into lower-rank components, tensor decomposition techniques can capture shared information across heterogeneous images and facilitate the discovery of common characteristics [6].

Variability: Tensor decomposition algorithms can effectively handle the variability present in cancer images by extracting robust and discriminative features. These algorithms have the ability to capture essential information while minimizing the influence of noise, artifacts, and imaging variations. Moreover, tensor decomposition methods can be utilized for denoising and artifact removal in cancer images, thereby improving their overall quality and consistency [6].

Integration of multi-modal data: Cancer studies frequently require the integration of information from various imaging modalities, including MRI, CT, and PET. Tensor decompositions offer a powerful framework for effectively fusing multi-modal data by modeling the joint relationships among different modalities. By decomposing the multi-modal tensor, tensor decomposition techniques can unveil shared and complementary information, facilitating a comprehensive analysis of multi-modal cancer images [15].

Tensor decomposition methods play a valuable role in the segmentation and annotation of cancer images. By decomposing the tensor into low-rank components, these methods can effectively separate different tissue types, tumors, or anatomical structures, thereby aiding in automated or semi-automated segmentation tasks. Furthermore, tensor decomposition techniques can be synergistically combined with other image analysis algorithms to enhance the accuracy and efficiency of the segmentation process [16].

Data reduction and efficiency: Cancer image datasets can be large and computationally demanding. Tensor decomposition algorithms can reduce the dimensionality of the data by capturing the most relevant information in a compact representation. This data reduction can lead to more efficient storage, retrieval, and analysis, enabling faster processing of large-scale cancer image datasets [12].

Interpretability and feature extraction: Tensor decompositions can extract interpretable features from cancer images, aiding in the understanding of underlying biological processes and disease characteristics. By decomposing the tensor into lower-rank components, tensor decomposition techniques can identify meaningful patterns and biomarkers in cancer images, providing insights into disease progression, treatment response, and prognosis [17].

By leveraging tensor decompositions, researchers can effectively address these challenges in cancer image analysis, leading to improved understanding, diagnosis, and treatment of cancer.

3.2 Omics Data

Omics data refers to large-scale molecular datasets that encompass various aspects of cellular activity and composition. These datasets are generated using high-throughput methods and provide comprehensive information. In the context of cancer research, omics data, including genomics, transcriptomics, and proteomics, play a crucial role in identifying molecular biomarkers associated with cancer development, progression, and treatment response. However, the challenge lies in merging disparate omics data types into a unified data structure and analyzing them collectively, posing additional difficulties [18]. Managing and analyzing such large-scale datasets requires the utilization of advanced computational and statistical methods to extract meaningful information. One of the significant challenges is integrating multiple omics datasets. In tensor decomposition, it has been found as a useful data analysis tool which provides an efficient way to integrate epigenomic data [9] [15]. In omics data, artificial intelligence (AI) plays a big role in the detection and monitoring of tumors. These AI tools can be used to decrease oversights and can be used as a detector against errors [14]. The integration of different types of omics data is crucial for obtaining a comprehensive understanding of cancer biology. However, this integration task is complex due to inherent biases, measurement errors, and differing scales inherent in these heterogeneous datasets [19].

Tensor decompositions can contribute to addressing various challenges in omics data analysis. Here are some specific challenges that tensor decompositions can help overcome in the context of omics data:

High dimensionality: Omics data, such as genomics, transcriptomics, and proteomics, are characterized by high-dimensional feature spaces. Tensor decompositions offer a powerful approach for dimensionality reduction by capturing the inherent low-rank structure in the data. By decomposing the omics tensor, tensor decomposition methods can effectively extract meaningful and concise representations of the data, facilitating downstream analysis [19].

Data integration: Integrating multiple omics datasets is crucial for a comprehensive analysis and understanding of complex biological systems. Tensor decompositions provide a natural framework for integrating heterogeneous omics data types. By decomposing the multi-omics tensor, tensor decomposition techniques can reveal shared and specific patterns across different data sources, enabling the identification of cross-modal relationships and uncovering novel insights [20].

Missing data imputation: Omics datasets often suffer from missing values due to experimental limitations or technical factors. Tensor decomposition methods can handle missing data by exploiting the latent structure in the tensor. By decomposing the tensor, tensor decomposition techniques can predict missing values based on the observed patterns, enabling the imputation of missing omics data and facilitating a more complete and reliable analysis [19].

Feature selection and extraction: Tensor decompositions can aid in feature selection and extraction from omics data. By decomposing the tensor into lower-rank components, tensor decomposition techniques can identify the most informative features or patterns associated with specific biological processes or phenotypes. This enables dimensionality reduction and the identification of relevant biomarkers, facilitating efficient and interpretable analysis of omics data [2].

Network analysis and pathway identification: Omics data often involve interactions and relationships between different biological entities, such as genes, proteins, and metabolites. Tensor decompositions can uncover network structures and pathways by capturing higher-order relationships in the data. By decomposing the tensor, tensor decomposition methods can identify co-regulated or co-occurring features, facilitating the identification of functional modules, pathway analysis, and biological interpretation.

Temporal and spatial analysis: Omics data collected over time or across spatial locations require methods that can capture temporal or spatial patterns [21]. Tensor decompositions can model the spatiotemporal structure in omics data by considering multiple dimensions simultaneously. By decomposing the tensor, tensor decomposition techniques can extract latent temporal or spatial factors, enabling the analysis of dynamic or spatially varying processes in omics data.

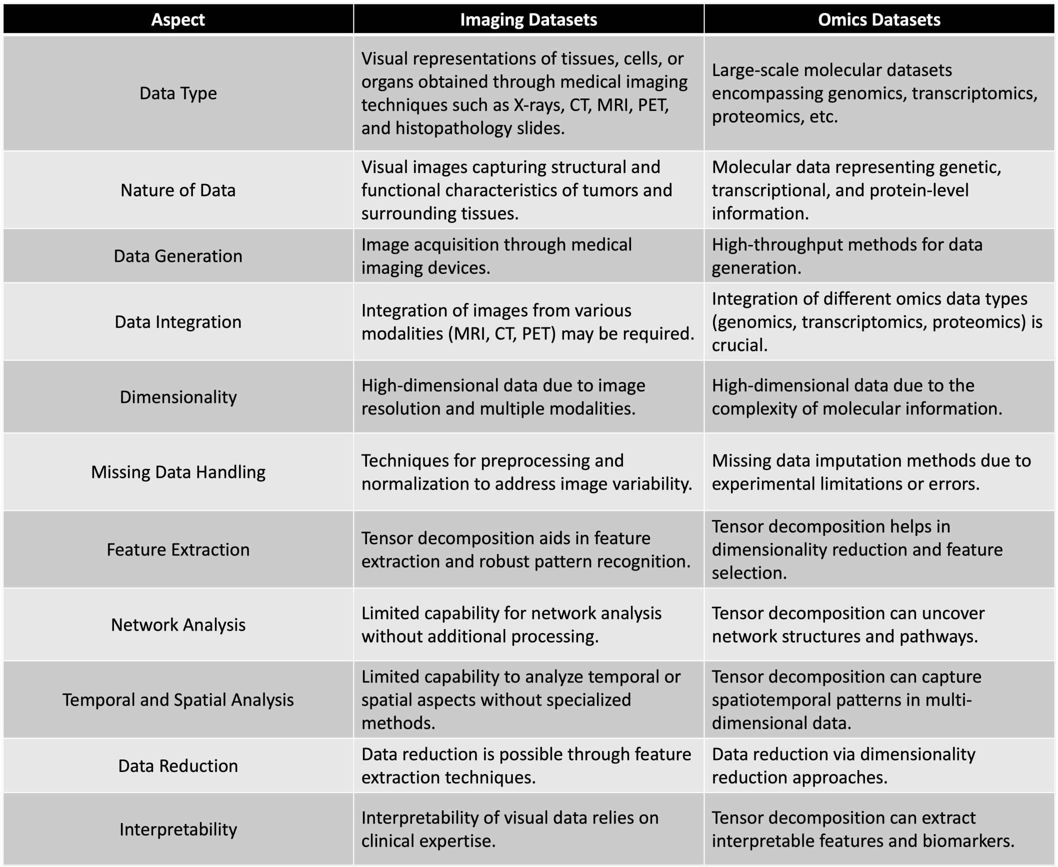
By leveraging tensor decompositions, researchers can address these challenges in omics data analysis, leading to a deeper understanding of biological systems, identification of biomarkers, and the discovery of novel insights in various omics domains.

4 Tensor Decomposition Methods

In this section of this paper, we will go over the approaches and methodologies that are used in the reviewed literature on cancer study using different tensor decomposition methods.

4.1 Canonical Polyadic (CP)

PARAFAC is a commonly employed tensor decomposition technique extensively utilized for cancer studies in image and multi-omics datasets. It serves as a method to break down multidimensional arrays, enabling focused analysis on specific aspects of interest and delivering a clear depiction of the obtained results. In the first equation (1), which represents a PARAFAC



Aspect	Imaging Datasets	Omics Datasets
Data Type	Visual representations of tissues, cells, or organs obtained through medical imaging techniques such as X-rays, CT, MRI, PET, and histopathology slides.	Large-scale molecular datasets encompassing genomics, transcriptomics, proteomics, etc.
Nature of Data	Visual images capturing structural and functional characteristics of tumors and surrounding tissues.	Molecular data representing genetic, transcriptional, and protein-level information.
Data Generation	Image acquisition through medical imaging devices.	High-throughput methods for data generation.
Data Integration	Integration of images from various modalities (MRI, CT, PET) may be required.	Integration of different omics data types (genomics, transcriptomics, proteomics) is crucial.
Dimensionality	High-dimensional data due to image resolution and multiple modalities.	High-dimensional data due to the complexity of molecular information.
Missing Data Handling	Techniques for preprocessing and normalization to address image variability.	Missing data imputation methods due to experimental limitations or errors.
Feature Extraction	Tensor decomposition aids in feature extraction and robust pattern recognition.	Tensor decomposition helps in dimensionality reduction and feature selection.
Network Analysis	Limited capability for network analysis without additional processing.	Tensor decomposition can uncover network structures and pathways.
Temporal and Spatial Analysis	Limited capability to analyze temporal or spatial aspects without specialized methods.	Tensor decomposition can capture spatiotemporal patterns in multi-dimensional data.
Data Reduction	Data reduction is possible through feature extraction techniques.	Data reduction via dimensionality reduction approaches.
Interpretability	Interpretability of visual data relies on clinical expertise.	Tensor decomposition can extract interpretable features and biomarkers.

Figure 2: Contrasting Imaging and Omics Datasets in Cancer Research

model, the variables we used hold different meanings/interpretations [17]. The f represents the number of components and it is used to define loading matrices A, B, and C of the dimensions I*F , J*F, and K*F, and this will use elements a_{if}, b_{if}, c_{kf} and the modeling error e_{ijk}.

$$x_{ijk} = \sum a_{if} b_{jf} c_{kf} + e_{ijk} \tag{1}$$

The article, [20] proposes a novel approach for predicting the progression of Alzheimer's disease using MRI (Magnetic Resonance Imaging) data. The method leverages tensor multi-task learning, which is a machine learning technique that jointly learns multiple related tasks to improve performance.The authors focus on capturing both spatial and temporal characteristics of the MRI data. They employ a tensor-based representation to model the multi-dimensional structure of the data and incorporate spatio-temporal similarity measurements. This approach enables the identification of patterns and changes in the MRI data over time, which are important indicators of Alzheimer's disease progression [16].

The objective of CP decomposition is to determine the factor matrices and their corresponding weights that can provide the closest approximation to the original tensor. This is usually accomplished by minimizing the sum of squared deviations between the original tensor and its estimated approximation [22]. According to reference [20], the figure 3 shows that

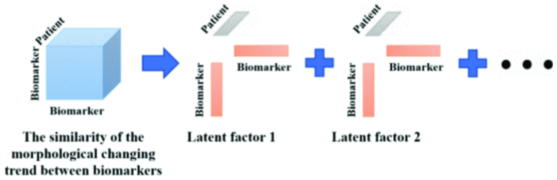


Figure 3: A graphical illustration of a PARAFAC model.

the PARAFAC decomposition process reveals that the morphological changing patterns among biomarkers, which indicate similarity, can be represented as multiple rank-one tensors [16]. Each rank-one tensor is obtained by taking the outer product of three latent components. The dimensions associated with the first biomarker, second biomarker, and patient samples are utilized to define each latent factor. Consequently, this approach provides an interpretable method for characterizing the latent factors responsible for the variability in the data [16]. In figure 3, A tensor is used to represent CP decomposition, which captures the resemblance in the morphological changing trend among biomarkers [16]. This uses the performed method of CP decomposition and extracts the set of rank-one latent factors from the original data. Another example of PARAFAC that is used is in reference [1]. This article talks about correlation tensor decomposition aiming to capture the correlations or relationships between voxels in a high-dimensional tensor. This technique can be useful for tasks such as denoising, dimensionality reduction, feature extraction, and data visualization [1].

4.2 TUCKER Decomposition

Tucker decomposition, also known as higher-order singular value decomposition (HOSVD), is a tensor factorization method that aims to decompose a tensor into a core tensor and factor matrices along each mode. It is an extension of the matrix singular value decomposition (SVD) to higher-dimensional data [20]. Due to its versatility and the manner in which it treats the PARAFAC model as a specific case, the Tucker model is commonly employed for decomposition, compression, and interpretation in various applications. According to reference [2], the study considered the data as a multi-way array, and that their approach is arranged in a 4-way tensors, and a Tucker decomposition with constraints that find the set of basis matrices and features for data [9]. Another way that the Tucker decomposition method is used, according to reference [22], is that it is used to fuse image information and gene information to gather features that can express a relationship between two modalities. This study focused on the treatment of non-small cell lung cancer. The researchers extracted 2D images from 3D data and preserved the spatial correlation of the original texture and edge information. This approach expanded the sample size of the image data to facilitate subsequent prediction [23]. According to reference [7], the researchers utilize an epigenomic tensor, a multidimensional representation of epigenetic data, to analyze complex biological datasets. This tensor captures information about DNA methylation patterns, histone modifications, and gene expression levels. The epigenomic tensor analysis incorporates epigenomic data, which focuses on the study of chemical modifications to DNA and histones that regulate gene expression. Epigenomic data is an important component of omics data and provides valuable insights into the functional elements of the genome [7].

4.3 Hierarchical Tensor Decomposition

A variation of the Tucker decomposition that focuses primarily on hierarchical tensor structures is called the Hierarchical Tucker Decomposition (HTD). Despite offering a flexible framework for high-dimensional data analysis, the Tucker decomposition may not be effective when dealing with large-scale tensors or tensors with hierarchical relationships. The HTD algorithm aims to capture the hierarchical structure within a tensor by representing it in a hierarchical form, where tensors at different levels of hierarchy correspond to different levels of detail or resolution. It allows for a more compact representation of hierarchical data, reducing the computational complexity and memory requirements compared to traditional Tucker decomposition [24]. In HTD, a tensor is represented as a tree-like structure, where each node corresponds to a tensor at a specific hierarchical level. The leaf nodes of the tree represent the original data tensor, while the internal nodes represent intermediate tensors obtained through hierarchical transformations. The root node corresponds to the highest level of hierarchy, representing the overall tensor [25]. In reference [24], the researchers showed how they could exploit tensor parallelism using HT tensors. They also wanted to show the potential of GPU tensor cores and utilize primitives to provide high performance [24].Hierarchical Tucker Decomposition (HTD) provides numerous benefits for analyzing hierarchical tensor structures. Firstly, HTD reduces computational complexity compared to traditional Tucker decomposition, enabling efficient analysis of large-scale tensors. By leveraging hierarchical dependencies, the algorithm generates a more compact representation, minimizing memory requirements and enhancing scalability [26]. HTD effectively captures both global and local dependencies within hierarchical data, facilitating feature extraction and analysis in diverse domains such as image and video processing, genomics, neuroscience, and social network analysis. Nevertheless, HTD encounters certain challenges [27]. Constructing an accurate hierarchical tree that accurately reflects underlying dependencies can be difficult, particularly when the hierarchical structure lacks clear definition or exhibits variation across different data instances. Scalability poses a concern for extremely large-scale tensors, as computational and memory demands escalate with tensor size and hierarchical depth. Interpreting the obtained hierarchical factors can be intricate, especially with deep hierarchies or when hierarchical levels entail diverse interpretations. Ongoing research and developments strive to address these challenges by exploring adaptive hierarchical structures and incorporating sparsity, thus augmenting the capabilities of hierarchical tensor decomposition methods [13] [28].

Figure 4 illustrates the number of published papers from 2013 to 2023 that used above mentioned methodologies.The choice between using CP decompositions and Tucker models in cancer studies depends on several factors, including the specific research goals, the nature of the data, and the research context. While CP decompositions and Tucker models are both tensor decomposition techniques, they have different strengths and weaknesses that make them suitable for different types of analyses. CP decomposition is commonly preferred in cancer studies for several reasons such as interpretability, simplicity, computational efficiency, and overfitting prevention. However, it is important to note that the choice between CP decompositions and Tucker models ultimately depends on the specific research objectives and the characteristics of the dataset. In certain scenarios, Tucker models may be more appropriate, such as when there is a need to model complex interactions or when the data exhibit significant dependencies across different modes.

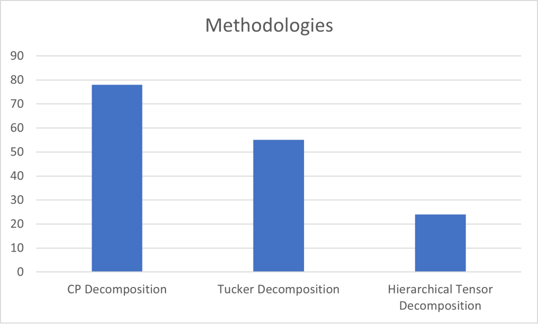


Figure 4: Number of published papers (2013-2023) for each methodology

5 Future Research Directions

The potential of tensor decompositions in advancing cancer studies through multi-dimensional data analysis is highlighted in this research. By addressing challenges and exploring novel applications, tensor-based approaches can improve cancer diagnosis, personalized treatment strategies, and our understanding of cancer biology.

One key research direction is the integration of multi-omics data, such as genomics, transcriptomics, epigenomics, and proteomics, using tensor decompositions. This holistic approach can uncover comprehensive molecular signatures associated with cancer subtypes, treatment response, and patient outcomes. Another direction involves incorporating spatial and temporal dimensions into tensor decompositions to analyze tumor evolution, genomic alterations, and dynamic changes in molecular profiles. Understanding the spatiotemporal dynamics of tumor growth, metastasis, and therapeutic resistance is crucial. Tensor decompositions can also be adapted for single-cell analysis, allowing for the identification of rare cell populations, characterizing cell-state transitions, and deciphering cellular networks within the tumor microenvironment. This approach provides insights into tumor heterogeneity and cellular dynamics. In addition to molecular data, integrating clinical variables alongside molecular data using tensor decompositions facilitates the identification of prognostic factors, personalized treatment strategies, and the development of predictive models for treatment response and survival outcomes. Although tensor decompositions offer data compression and feature extraction, interpreting the extracted latent factors can be challenging.

Future research should focus on developing interpretable tensor decomposition methods that provide meaningful insights and visualizations, aiding in the understanding of complex molecular interactions, biological processes, and clinical implications. Addressing missing values in tensors, especially in large-scale omics datasets, is important for accurate analysis. Developing tensor-based methods for missing value imputation can enhance data quality and enable analysis of complete datasets. Furthermore, bridging the gap between Tucker3 decomposition and chemometrics in cancer research can unlock new avenues for understanding the molecular underpinnings of cancer, identifying biomarkers, and developing predictive models. Applying Tucker3 decomposition in chemometrics for cancer studies can enhance our understanding of cancer biology at a molecular level, leading to improved patient outcomes and personalized medicine approaches.

6 Conclusion

Tensor decompositions have emerged as a valuable tool in the field of cancer study, offering powerful techniques for analyzing complex and high-dimensional datasets. By representing multi-dimensional data as tensors, these methods enable researchers to extract meaningful patterns, identify relevant biomarkers, and gain deeper insights into the underlying mechanisms of cancer development, progression, and treatment.The integration of tensor decomposition techniques with other analytical approaches promises to revolutionize cancer research and pave the way for improved diagnostics, treatments, and outcomes for patients worldwide. This paper provides an extensive overview of research conducted between 2013 and 2023, specifically examining the recent progress made in cancer studies through the application of tensor decompositions. By incorporating imaging data and other multi-omics information, tensor decomposition algorithms have emerged as highly influential tools in cancer research. The complexities stemming from diverse data sources have highlighted the importance of these algorithms in identifying biomarkers, emphasizing the challenges, limitations, and potential advancements in this field.

7 Acknowledgements

This work is supported in part by grant number 2050883 from the National Science Foundation.

References

[1] Y. Deng, X. Tang, and A. Qu, "Correlation tensor decomposition and its application in spatial imaging data," *Journal of the American Statistical Association*, vol. 118, no. 541, pp. 440–456, 2023.

[2] S. Hussain, I. Mubeen, N. Ullah, S. S. U. D. Shah, B. A. Khan, M. Zahoor, R. Ullah, F. A. Khan, and M. A. Sultan, "Modern diagnostic imaging technique applications and risk factors in the medical field: A review," Jun 2022.

[3] X. Tang, X. Bi, and A. Qu, "Individualized multilayer tensor learning with an application in imaging analysis," *Journal of the American Statistical Association*, vol. 115, no. 530, pp. 836–851, 2020.

[4] N. Perakakis, A. Yazdani, G. E. Karniadakis, and C. Mantzoros, "Omics, big data and machine learning as tools to propel understanding of biological mechanisms and to discover novel diagnostics and therapeutics," *Metabolism*, vol. 87, pp. A1–A9, 2018.

[5] D. Diaz, A. Bollig-Fischer, and A. Kotov, "Tensor decomposition for sub-typing of complex diseases based on clinical and genomic data," in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 647–651, 2019.

[6] E. López de Maturana, L. Alonso, P. Alarcón, I. A. Martín-Antoniano, S. Pineda, L. Piorno, M. L. Calle, and N. Malats, "Challenges in the integration of omics and non-omics data," *Genes*, vol. 10, no. 3, 2019.

[7] J. R. Leistico, P. Saini, C. R. Futtner, M. Hejna, Y. Omura, P. N. Soni, P. Sandlesh, M. Milad, J.-J. Wei, S. Bulun, J. B. Parker, G. D. Barish, J. S. Song, and D. Chakravarti, "Epigenomic tensor predicts disease subtypes and reveals constrained tumor evolution," *Cell Reports*, vol. 34, no. 13, p. 108927, 2021.

[8] Q. Qiao, Y.-L. Gao, S.-S. Yuan, and J.-X. Liu, "Robust tensor method based on correntropy and tensor singular value decomposition for cancer genomics data," in *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 509–514, 2021.

[9] A. Fargeas, L. Albera, A. Kachenoura, G. Dréan, J.-D. Ospina, J. Coloigner, C. Lafond, J.-B. Delobel, R. De Crevoisier, and O. Acosta, "On feature extraction and classification in prostate

cancer radiotherapy using tensor decompositions," *Medical Engineering Physics*, vol. 37, no. 1, pp. 126–131, 2015.

[10] F. Koleini, S. Hugelier, M. A. Lakeh, H. Abdollahi, J. Camacho, and P. J. Gemperline, "On the complementary nature of anova simultaneous component analysis (asca+) and tucker3 tensor decompositions on designed multi-way datasets," *Journal of Chemometrics*, p. e3514, 2023.

[11] X. Luciani and L. Albera, "Canonical polyadic decomposition based on joint eigenvalue decomposition," *Chemometrics and Intelligent Laboratory Systems*, vol. 132, pp. 152–167, 2014.

[12] T. D. Pham, V. Ravi, C. Fan, B. Luo, and X.-F. Sun, "Tensor decomposition of largest convolutional eigenvalues reveals pathologic predictive power of rhob in rectal cancer biopsy," *The American Journal of Pathology*, vol. 193, no. 5, pp. 579–590, 2023.

[13] J. Xue, Y. Zhao, S. Huang, W. Liao, J. C.-W. Chan, and S. G. Kong, "Multilayer sparsity-based tensor decomposition for low-rank tensor completion," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 11, pp. 6916–6930, 2022.

[14] W. L. Bi, A. Hosny, M. B. Schabath, M. L. Giger, N. J. Birkbak, A. Mehrtash, T. Allison, O. Arnaout, C. Abbosh, I. F. Dunn, R. H. Mak, R. M. Tamimi, C. M. Tempany, C. Swanton, U. Hoffmann, L. H. Schwartz, R. J. Gillies, R. Y. Huang, and H. J. W. L. Aerts, "Artificial intelligence in cancer imaging: Clinical challenges and applications," *CA: A Cancer Journal for Clinicians*, vol. 69, no. 2, pp. 127–157, 2019.

[15] G. Lu, L. Halig, D. Wang, Z. G. Chen, and B. Fei, "Spectral-spatial classification using tensor modeling for cancer detection with hyperspectral imaging," in *Medical Imaging 2014: Image Processing* (S. Ourselin and M. A. Styner, eds.), vol. 9034, p. 903413, International Society for Optics and Photonics, SPIE, 2014.

[16] Y. Zhang, P. Yang, and V. Lanfranchi, "Tensor multi-task learning for predicting alzheimer's disease progression using mri data with spatio-temporal similarity measurement," in *2021 IEEE 19th International Conference on Industrial Informatics (INDIN)*, pp. 1–8, 2021.

[17] F. Koleini, P. Gemperline, and N. Tabrizi, "Biomarker discovery in multi-omics datasets using tensor decompositions; a comprehensive review," in *Proceedings of International Conference on Bioinformatics and Computational Biology (BICOB-2023)* (H. Al-Mubaid, T. Aldwairi, and O. Eulenstein, eds.), vol. 92 of *EPiC Series in Computing*, pp. 11–24, EasyChair, 2023.

[18] I. Jung, M. Kim, S. Rhee, S. Lim, and S. Kim, "Monti: A multi-omics non-negative tensor decomposition framework for gene-level integrative analysis," *Frontiers in Genetics*, vol. 12, 2021.

[19] B. Palsson and K. Zengler, "The challenges of integrating multi-omic data sets," Oct 2010.

[20] Y.-h. Taguchi and T. Turki, "Tensor-decomposition-based unsupervised feature extraction applied to prostate cancer multiomics data," *Genes*, vol. 11, p. 1493, Dec 2020.

[21] S.-N. Zhang, Y.-L. Zhang, J.-X. Liu, J. Wang, J. Shang, and D.-H. Ge, "Tensor robust pca based on transformed tensor singular value decomposition for cancer genomic data," in *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 2162–2168, 2022.

[22] N. B. Erichson, K. Manohar, S. L. Brunton, and J. N. Kutz, "Randomized cp tensor decomposition," *Machine Learning: Science and Technology*, vol. 1, p. 025012, may 2020.

[23] G. Hou, L. Jia, Y. Zhang, W. Wu, L. Zhao, J. Zhao, L. Wang, and Y. Qiang, "Deep learning approach for predicting lymph node metastasis in non-small cell lung cancer by fusing image–gene data," *Engineering Applications of Artificial Intelligence*, vol. 122, p. 106140, 2023.

[24] H. Huang, X.-Y. Liu, W. Tong, T. Zhang, A. Walid, and X. Wang, "High performance hierarchical tucker tensor learning using gpu tensor cores," *IEEE Transactions on Computers*, vol. 72, no. 2, pp. 452–465, 2023.

[25] A.-H. Phan, A. Cichocki, A. Uschmajew, P. Tichavský, G. Luta, and D. P. Mandic, "Tensor networks for latent variable analysis: Novel algorithms for tensor train approximation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 11, pp. 4622–4636, 2020.

[26] J. Vendrow, J. Haddock, and D. Needell, "A generalized hierarchical nonnegative tensor decomposition," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal*

Processing (ICASSP), pp. 4473–4477, 2022.

[27] A.-H. Phan, P. Tichavský, and A. Cichocki, “Partitioned hierarchical alternating least squares algorithm for cp tensor decomposition,” in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2542–2546, 2017.

[28] R. Zdunek and K. Fonal, “Randomized nonnegative tensor factorization for feature extraction from high-dimensional signals,” in *2018 25th International Conference on Systems, Signals and Image Processing (IWSSIP)*, pp. 1–5, 2018.