

# Study on Breast Cancer Prediction and Diagnosis of MRI Image using Machine Learning Technique

Alka chouhan<sup>1</sup>, Swati khanve<sup>2</sup>, Nitya khare<sup>3</sup>

<sup>1</sup>*M. Tech. Scholar, Dept. of Computer Science and Engineering, Sagar Institute of Research Technology Excellence, Bhopal, India*

<sup>2</sup>*Assistant Professor, Dept. of Computer Science and Engineering, Sagar Institute of Research Technology Excellence, Bhopal, India*

<sup>3</sup>*Assistant Professor & HOD, Dept. of Computer Science and Engineering, Sagar Institute of Research Technology Excellence, Bhopal, India*

**Abstract:** - Breast cancer is one of the most prevalent and life-threatening diseases among women worldwide, where early and accurate diagnosis is crucial for improving patient outcomes. Magnetic Resonance Imaging (MRI) is a highly sensitive imaging modality for breast cancer detection; however, manual interpretation of MRI scans is time-consuming and prone to inter-observer variability. In recent years, machine learning (ML) and deep learning (DL) techniques have been extensively explored to automate breast cancer prediction and diagnosis using MRI images. This paper presents a comprehensive and structured review of MRI-based breast cancer diagnosis techniques employing machine learning approaches. The reviewed studies are systematically categorized into traditional machine learning methods, radiomics- based models, and deep learning approaches, including convolutional neural networks and transfer learning frameworks. A critical comparative analysis is performed by summarizing key quantitative outcomes such as accuracy, sensitivity, specificity, datasets, and evaluation metrics. The review highlights that deep learning models consistently outperform traditional machine learning techniques due to their ability to automatically extract discriminative features from MRI data. However, challenges such as limited dataset size, lack of standardized benchmarking, high computational complexity, and limited model interpretability remain significant barriers to clinical adoption.

Finally, this survey identifies existing research gaps and outlines future research directions, emphasizing the need for large-scale standardized MRI datasets, explainable artificial intelligence models, and robust evaluation frameworks. The findings of this review aim to assist researchers and clinicians in developing reliable and clinically applicable MRI- based breast cancer prediction systems.

**Keywords:** - MRI Image, Machine Learning, Cancer

## I. INTRODUCTION

Cancer is a disease of the genes involved in developmental, cell growth, and differentiation processes. According to the global study, an estimated 8.2 million cancer deaths and 14.1 million new cancer cases are reported in 2012 worldwide [1]. The number of estimated cancer deaths and new cancer cases are increased to 9.6 million and 18.1 million in 2018 globally [2]. Among the men and women, lung and breast cancers are the most frequently diagnosed and the leading cause of cancer death, respectively. Other frequently diagnosed cancers include prostate, liver, stomach, colorectum and cervix uteri. The report [1] suggests developed countries are mainly hit by bladder cancer and uterine cancer among males and females. In the case of less developed countries, liver and stomach cancer are the leading causes of cancer death. Less developed countries account for only 65% of cancer deaths worldwide, in spite of their larger population. However, the number of cancer cases is increasing in underdeveloped countries due to massive population growth and the increasing prevalence of the known risk factors [3]. It is needless to say that cancer is a global health concern. Better diagnoses, treatment as well as an understanding of molecular mechanisms are needed to overcome cancer- related problems. Healthy cells grow and divide, when stimulated by growth factors but have many controls on that growth. In the case of damage cells, a molecular brake stops them from dividing until they are repaired. If they can't be fixed, they commit programmed cell death, called apoptosis. Cancer cells have defects in their control mechanisms that govern how often they will divide. The reason behind this is the proteins that

control them are altered, leading to increased cell division within the tumor [4, 5]. Typically, normal cells of the body require hormones and other molecules that act as signals for them to grow and divide. However, cancer cells can grow without these external signals by permanently activating the signaling pathways. Cancer is a collection of complex miscellaneous diseases with known genetic risk factors. Recent studies have found, microRNA (miRNA) play an essential role in gene expression regulation by binding to the 3' untranslated regions (UTRs) of target messenger RNAs (mRNAs), causing a block of translation and/or messenger RNA (mRNA) degradation [6]. Variations in the level of expression of distinct miRNAs have been observed in the development and progression of multiple human cancers and can arise from either genetic or epigenetic means. Evidence suggests that miRNAs function as oncogenes or tumor suppressors [7–9]. Thus, changes in miRNA expression may promote carcinogenesis by regulating the expression patterns of vital genes involved in tumor growth and progression [10].

1.1 Breast Cancer

Breast Cancer is a serious and widespread disease caused by the uncontrolled growth of abnormal cells in breast tissue. It is one of the most common cancers affecting women worldwide, although it can also

occur in men in rare cases. Early detection of breast cancer significantly increases survival rates and improves treatment outcomes [11].

Breast cancer is broadly classified into several types, including ductal carcinoma, lobular carcinoma, and invasive breast cancer. Common symptoms include the presence of a lump in the breast, changes in breast size or shape, skin dimpling, nipple discharge, and localized pain. However, in the early stages, breast cancer may not present noticeable symptoms, making diagnosis more challenging.

Diagnostic procedures typically involve imaging techniques such as mammography, ultrasound, and Magnetic Resonance Imaging (MRI). Among these, MRI offers high sensitivity and is particularly useful for detecting early-stage and complex breast lesions. In recent years, machine learning and deep learning techniques have gained significant attention in breast cancer diagnosis by enabling automated analysis of medical images. These techniques can extract meaningful features from MRI images and accurately classify tumors as benign or malignant [12].

Overall, the integration of advanced imaging modalities with machine learning-based computer-aided diagnosis systems has shown great potential in improving diagnostic accuracy, reducing human error, and supporting clinicians in early breast cancer detection and decision-making processes.

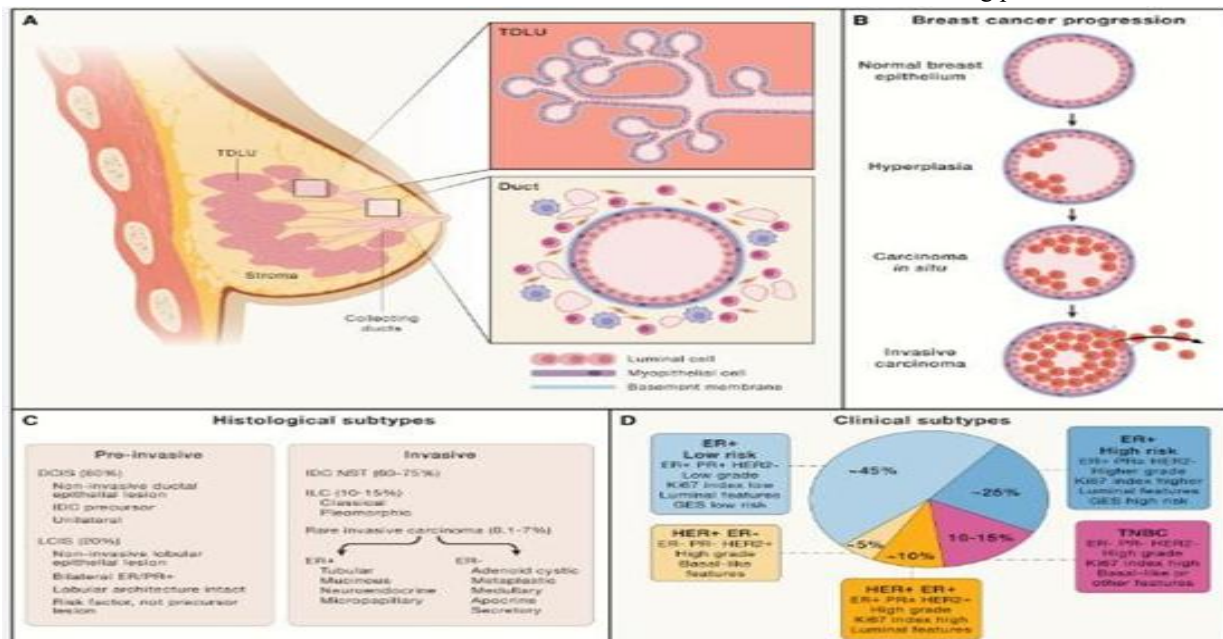


Figure 1: Breast Cancer process

II. LITERATURE REVIEW

Early research on breast cancer prediction primarily

employed traditional machine learning (ML) algorithms combined with handcrafted features extracted from MRI images. Bista *et al.* [1] proposed a breast cancer prediction system using classical ML classifiers such as Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbor (KNN). Their study demonstrated that ML models can effectively distinguish malignant and benign cases; however, the performance strongly depended on feature engineering and preprocessing techniques. While these approaches are computationally efficient, their reliance on manually designed features limits robustness and generalization across diverse MRI datasets.

Radiomics-based approaches have gained attention due to their ability to extract high-dimensional quantitative features from MRI images. Matsuda *et al.* [2] evaluated the diagnostic accuracy of a radiomics-driven ML model using synthetic breast MRI. The study reported improved classification performance by integrating texture, shape, and intensity-based features. Despite promising results, the authors emphasized that radiomics models require careful feature selection and standardization, and their performance may degrade when applied to external datasets. The dependence on handcrafted radiomic features introduces challenges related to reproducibility and cross-institutional variability, highlighting the need for more automated and scalable solutions.

Recent systematic reviews provide a comprehensive evaluation of MRI-based ML and DL models. Zhang *et al.* [3] conducted a meta-analysis assessing diagnostic accuracy across multiple MRI-based ML studies and reported consistently high sensitivity and specificity for deep learning models compared to traditional ML approaches. Similarly, Lee *et al.* [4] reviewed deep learning-based breast MRI techniques for predicting axillary lymph node metastasis, emphasizing the clinical potential of CNN-based models while noting concerns regarding interpretability and clinical deployment.

Liang *et al.* [5] developed an MRI-based deep learning model to predict lymphovascular invasion in node-negative invasive breast cancer. Their findings

confirmed the strong predictive capability of deep models, although the study noted the requirement for large annotated datasets and high computational resources.

Houssami *et al.* [6] reviewed current applications and future directions of machine learning in breast MRI, highlighting challenges such as limited dataset diversity, lack of standardized evaluation metrics, and reduced model explainability.

Zheng *et al.* [7] conducted a comparative analysis of multiple ML classifiers, including SVM, Decision Trees, and Random Forest, on breast MRI images. Their results showed moderate to high classification accuracy, but the study also highlighted sensitivity to dataset imbalance and variability in MRI acquisition protocols. These limitations indicate that traditional ML methods may struggle to capture complex spatial patterns inherent in breast MRI data.

Deep learning (DL) techniques, particularly Convolutional Neural Networks (CNNs), have shown superior performance in breast MRI analysis by automatically learning hierarchical features from raw images. Zhang *et al.* [8] employed transfer learning using pretrained CNN architectures for breast cancer diagnosis on MRI data. Their approach significantly improved classification accuracy compared to traditional ML methods, demonstrating the effectiveness of feature transfer from large-scale natural image datasets.

Although existing studies demonstrate the effectiveness of machine learning and deep learning techniques for breast cancer diagnosis using MRI, several limitations remain. Most studies rely on small or institution-specific datasets, limiting generalizability. Furthermore, the lack of standardized performance metrics and insufficient comparative evaluations across ML and DL models hinder objective assessment. Deep learning approaches, while highly accurate, often function as black-box models, reducing clinical trust and interpretability. These gaps justify the need for a structured survey that systematically compares MRI-based ML techniques, datasets, and quantitative outcomes to guide future research.

Table 1: Comparative Summary of MRI-Based Breast Cancer Prediction Studies

Ref.	Year	MRI Type / Dataset	Methodology	Performance Metrics	Key Limitations
[1]	2024	Breast MRI (Private)	SVM, RF, KNN	Accuracy $\approx$ 91%	Manual feature extraction
[2]	2025	Synthetic Breast MRI	Radiomics + ML	High diagnostic accuracy (AUC > 0.90)	Feature dependency, dataset specificity
[3]	2025	Multiple MRI datasets	ML & DL (Meta-analysis)	Sensitivity & Specificity > 85%	Study heterogeneity
[4]	2025	Breast MRI	Deep CNN Models	Improved lymph node prediction accuracy	Limited interpretability
[5]	2024	Breast MRI	Deep Learning (CNN)	High predictive performance	Requires large labeled datasets
[6]	2021	Breast MRI (Survey)	ML & DL Review	Qualitative comparison	Lack of quantitative benchmarking
[7]	2021	Breast MRI	SVM, DT, RF	Moderate–High accuracy	Dataset imbalance
[8]	2019	Breast MRI	Transfer Learning CNN	Improved accuracy over ML	High computational cost

**Problem Identification:-**

Breast cancer remains one of the leading causes of cancer-related mortality among women worldwide, and early and accurate diagnosis plays a critical role in improving patient survival rates. Magnetic Resonance Imaging (MRI) is widely recognized as a highly sensitive imaging modality for breast cancer detection due to its superior soft-tissue contrast and ability to capture tumor heterogeneity. However, manual interpretation of breast MRI scans is time-consuming, subjective, and highly dependent on radiologist expertise, leading to inter-observer variability and diagnostic inconsistencies.

Although numerous machine learning and deep learning techniques have been proposed for breast cancer prediction and diagnosis using MRI images, existing studies suffer from several key limitations. First, many approaches rely on small, institution-specific datasets, which restricts the generalizability and robustness of the developed models. Second, there is a lack of standardized evaluation metrics and benchmarking protocols, making it difficult to objectively compare the performance of different machine learning methods. Third, traditional machine

learning models depend heavily on handcrafted features, while deep learning approaches, despite achieving higher accuracy, often function as black-box models with limited interpretability, hindering clinical adoption.

Furthermore, existing literature lacks a comprehensive and structured survey that critically compares traditional machine learning, radiomics-based, and deep learning approaches specifically for MRI-based breast cancer diagnosis. Most reviews either focus on non-imaging data or do not provide systematic quantitative comparisons of datasets, methodologies, and performance outcomes. These gaps highlight the need for a focused and structured review that synthesizes current MRI-based machine learning techniques, identifies their limitations, and outlines future research directions for clinically reliable breast cancer diagnosis.

**III. BREAST CANCER USING MACHINE LEARNING**

Breast cancer diagnosis using machine learning has emerged as a powerful approach to improve early detection, accuracy, and clinical decision-making.

Traditional diagnostic methods rely heavily on manual interpretation of medical images and clinical data, which can be time-consuming and subject to human error. Machine learning techniques enable automated analysis of large volumes of data to identify complex patterns that are often difficult to detect through conventional methods.

In machine learning-based breast cancer diagnosis, medical imaging modalities such as MRI, mammography, and ultrasound are commonly used. The process typically involves image preprocessing to reduce noise and enhance image quality, followed by segmentation to isolate the region of interest. Feature extraction techniques are then applied to obtain relevant texture, shape, and intensity features. These features are used to train supervised machine learning models such as Support Vector Machine (SVM), Random Forest (RF), k-Nearest Neighbor (k-NN),

Naïve Bayes, and Decision Trees for classifying breast tumors as benign or malignant.

Recent advancements in deep learning, particularly Convolutional Neural Networks (CNNs), have further enhanced diagnostic performance by automatically learning hierarchical features directly from images. Machine learning-based systems have demonstrated high accuracy, sensitivity, and specificity, helping reduce false diagnoses and improve early cancer detection.

Overall, the integration of machine learning with breast cancer diagnosis provides a reliable, efficient, and scalable solution for computer-aided diagnosis systems. These approaches assist radiologists by reducing workload, improving diagnostic consistency, and enabling timely treatment, thereby contributing to better patient outcomes in modern healthcare systems.

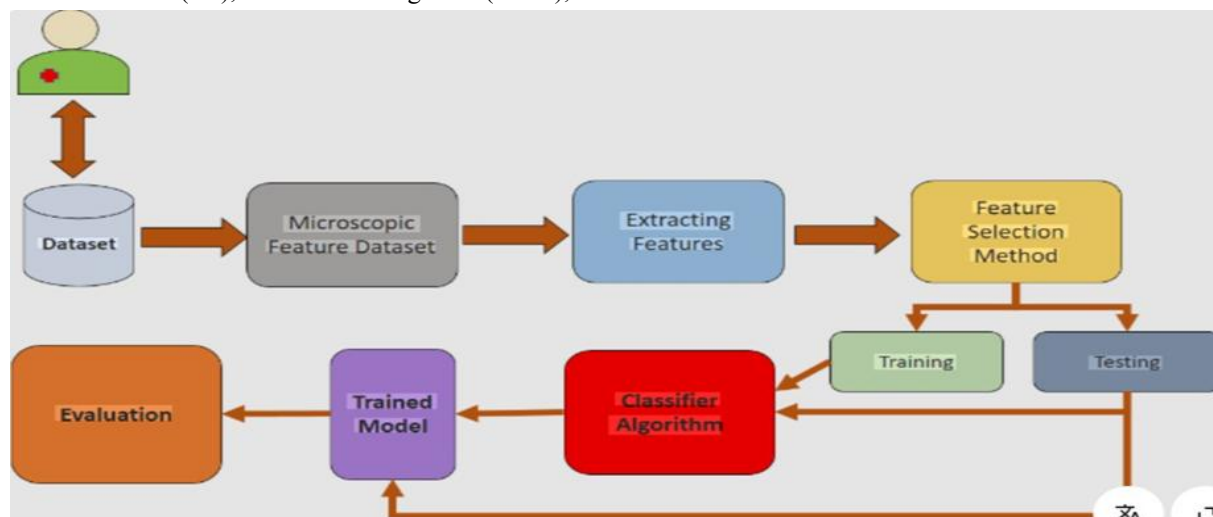


Figure 2: Process of ML

### Machine Learning

Machine Learning is a subset of Artificial Intelligence concerned with “teaching” computers how to act without being explicitly programmed for every possible scenario. The central concept in Machine Learning is developing algorithms that can self-learn by training on a massive number of inputs. Machine learning algorithms are used in various applications, such as email filtering and computer vision, where it is difficult or infeasible to develop conventional algorithms to perform the needed tasks [13]. Machine learning enables the analysis of vast amounts of information. While it usually delivers faster, more precise results to identify profitable prospects or dangerous risks, it may also require additional time

and assets to train it appropriately. Merging machine learning with AI and perceptive technologies can make it even more effective in processing vast volumes of information. Machine learning is closely associated with computational statistics, which focuses on making predictions using computers. Machine learning approaches are conventionally divided into three broad categories, namely Supervised Learning, Unsupervised Learning & Semi-supervised Learning, depending on the nature of the "signal" or "feedback" available to the learning system [14].

### Supervised Learning

A model is trained through a process of learning in

which predictions must be made and corrected if those predictions are wrong. The training process continues until a desired degree of accuracy is reached on the training data. Input data is called training data and has a known spam / not-spam label or result at one time.

#### Unsupervised Learning

By deducting the structures present in the input data, a model is prepared. This may be for general rules to be extracted. It may be through a mathematical process that redundancy can be systematically reduced, or similar data can be organized. There is no labeling of input data, and there is no known result.

#### Semi-Supervised Learning

Semi-supervised learning fell between unsupervised learning (without any labeled training data) and supervised learning (with completely labeled training data). There is a desired problem of prediction, but the model needs to learn the structures and make predictions to organize the data. Input data is a combination of instances that are marked and unlabeled.

#### IV. EXPECTED OUTCOME

The expected outcomes are follow:

- The proposed hybrid machine learning model will expect to accurately classify breast MRI images into benign and malignant categories.
- The model can be extended to larger datasets and integrated with clinical systems.
- The system is anticipated to achieve high performance in terms of accuracy, precision and recall.
- The system is expected to minimize false positives and false negatives compared to traditional diagnosis methods.

#### V. CONCLUSION

This review presented a comprehensive and structured analysis of machine learning and deep learning techniques for breast cancer prediction and diagnosis using MRI images. By systematically categorizing existing studies into traditional machine learning, radiomics-based methods, and deep learning approaches, the survey highlighted the evolution of MRI-based breast cancer diagnosis and critically compared their methodologies, datasets, and

quantitative performance outcomes. The comparative analysis demonstrated that deep learning models, particularly convolutional neural networks and transfer learning frameworks, consistently outperform traditional machine learning techniques by automatically learning discriminative features from MRI data.

Despite the promising diagnostic accuracy achieved by existing models, several challenges remain unresolved. Most studies rely on limited and non-standardized datasets, which restricts model generalizability and reproducibility. Additionally, the absence of unified benchmarking protocols and consistent evaluation metrics complicates fair performance comparison across different approaches. Furthermore, although deep learning techniques achieve superior accuracy, their lack of interpretability and high computational requirements pose significant barriers to real-world clinical deployment.

Overall, this survey identifies critical research gaps and emphasizes the need for large-scale, standardized MRI datasets, explainable machine learning models, and robust comparative evaluations to enhance clinical reliability. Future research should focus on integrating explainable artificial intelligence, multimodal MRI analysis, and clinically validated frameworks to bridge the gap between research advancements and practical breast cancer diagnosis. The findings of this review aim to serve as a valuable reference for researchers and clinicians working toward accurate, efficient, and trustworthy MRI-based breast cancer prediction systems.

#### REFERENCES

- [1] C. Bista, A. M, S. Slimanzay, M. S. Sheikh and P. Srinivasa Rao, "Breast Cancer Prediction System Utilizing Machine Learning Algorithms," *2024 IEEE AITU: Digital Generation*, Astana, Kazakhstan, 2024, pp. 80-84.
- [2] T. Matsuda, M. Matsuda, H. Haque *et al.*, "Diagnostic accuracy of a machine learning model using radiomics features from breast synthetic MRI," *BMC Med. Imaging*, vol. 25, art. no. 399, pp. 1–11, Sept. 2025.
- [3] J. Zhang, Q. Wu, P. Lei *et al.*, "Diagnostic accuracy of machine learning-based magnetic resonance imaging models in breast cancer classification: A systematic review and meta-

- analysis,” *World J. Surg. Onc.*, vol. 23, art. no. 231, pp. 1–13, Jun. 2025.
- [4] C. F. Lee, J. Lin, Y.-L. Huang *et al.*, “Deep learning-based breast MRI for predicting axillary lymph node metastasis: A systematic review and meta-analysis,” *Cancer Imaging*, vol. 25, art. no. 44, pp. 1–15, Mar. 2025.
- [5] R. Liang, F. Li, J. Yao *et al.*, “Predictive value of MRI-based deep learning model for lymphovascular invasion status in node-negative invasive breast cancer,” *Sci. Rep.*, vol. 14, art. no. 16204, pp. 1–12, Jul. 2024.
- [6] G. Houssami, M. Turner and R. Morrow, “Machine Learning for Breast MRI: Current Applications and Future Directions,” *European Radiology*, vol. 31, no. 6, pp. 3752–3764, 2021.
- [7] Y. Zheng, B. Liu and S. Chen, “Comparative Study of Machine Learning Techniques for Breast Cancer Classification Using MRI Images,” *Biomedical Signal Processing and Control*, vol. 68, Article ID 102645, 2021.
- [8] X. Zhang, Y. Chen and Z. Wang, “Breast Cancer Diagnosis Using Transfer Learning with Pretrained CNN Models on MRI,” *Computers in Biology and Medicine*, vol. 115, Article ID 103498, 2019.
- [9] Khurma RA, Aljarah I, Sharieh A, Elaziz MA, Damaševičius R, Krilavičius T. A review of the modification strategies of the nature inspired algorithms for feature selection problem. *Mathematics*. 2022;10(3):1–45.
- [10] Jain AK. Data clustering: 50 years beyond K-means. *Pattern Recogn Lett*. 2010;31(8):651–66.
- [11] Murtagh F, Contreras P. Algorithms for hierarchical clustering: an overview. *Wiley Interdiscip Rev Data Min Knowl Discov*. 2012;2(1):86–97.
- [12] Dabhi DP, Patel MR, Dipak MRP, Dabhi P. Extensive survey on hierarchical clustering methods in data mining. *Int Res J Eng Technol*, 2016; 03(11):659–665, [Online]. Available: <https://www.irjet.net/archives/V3/i11/IRJET-V3I11115.pdf>
- [13] Kriegel HP, Kröger P, Sander J, Zimek A. Density-based clustering. *Wiley Interdiscip Rev Data Min Knowl Discov*. 2011;1(3):231–40.
- [14] Moulavi D, Jaskowiak PA, Campello RJGB, Zimek A, Sander J. Density-based clustering validation. *SIAM Int Conf Data Min 2014, SDM 2014*. 2014; 2(i):839–847.
- [15] Aziz R, Verma CK, Srivastava N. A novel approach for dimension reduction of microarray. *Comput Biol Chem*. 2017;71:161–169.
- [16] Aziz R, Verma CK, Srivastava N. Dimension reduction methods for microarray data: a review. *AIMS Bioeng*. 2017;4(1):179–97.
- [17] Van Der Maaten LJP, Postma EO, Van Den Herik HJ. Dimensionality reduction: a comparative review. *J Mach Learn Res*. 2009; 10:1–41.