International Journal of Engineering Technology & Advanced Learning
ISSN: 3107-4227, Vol. 02 (01), March 2026, pp. 83-87

Deep Learning–Based Anomaly Detection for Cancer Analysis Using Autoencoder Models – A Survey

Anju Mehra[1], Irfan Khan[1], Damodar Prasad Tiwari[1], Kailash Patidar[1]

[1]*Department of Computer Engineering, Bansal Institute of Science and Technology, Bhopal, India*

Abstract

Cancer diagnosis remains a challenging task due to the high dimensionality of biomedical data, complex feature relationships, and severe class imbalance between normal and abnormal samples. With the rapid advancement of artificial intelligence, deep learning techniques have emerged as powerful tools for improving cancer detection and analysis. This research thoroughly examines deep learning-based anomaly detection techniques for cancer analysis, with an emphasis on diverse datasets and algorithmic frameworks. The survey examines preprocessing strategies, dataset characteristics, model architectures, and evaluation metrics used across different studies. The analysis highlights the strengths and limitations of existing approaches and identifies key challenges such as data imbalance, limited labeled data, and model interpretability. The paper also outlines potential research directions, including hybrid models, explainable artificial intelligence, and multimodal data integration, to enhance the effectiveness of cancer detection systems.

Keywords: Cancer Detection, Anomaly Detection, Deep Learning, Autoencoder, Reconstruction Error, Medical Data Analysis.

1. Introduction

Cancer remains one of the leading causes of mortality worldwide, posing significant challenges due to its complex biological characteristics and diverse manifestations. To increase patient survival rates and facilitate efficient treatment planning, early detection and precise diagnosis are crucial. However, managing large-scale and high-dimensional biomedical data frequently presents challenges for conventional diagnostic techniques. With the rapid growth of healthcare digitization, vast amounts of data are generated through medical imaging systems, electronic health records, and genomic sequencing. These datasets provide valuable insights but also introduce challenges related to data complexity, heterogeneity, and imbalance. In particular, cancer datasets are often characterized by a significant imbalance between normal and abnormal samples, which affects the performance of conventional classification models. To solve these issues, deep learning and machine learning approaches have been widely used. While deep learning models like convolutional neural networks allow autonomous feature extraction from raw data, traditional machine learning techniques like support vector machines and decision trees rely on feature engineering and labeled data.. More recently, anomaly detection approaches have gained attention as they focus on learning normal data patterns and identifying deviations, making them suitable for imbalanced datasets. This paper presents a comprehensive survey of different datasets and algorithms used in deep learning–based cancer analysis. The study focuses on evaluating various modeling approaches, including supervised and unsupervised techniques, with particular emphasis on autoencoder-based anomaly detection models. The objective is to provide a comparative understanding of existing methods and identify key challenges and future research directions in this domain.

Copyright ©IJETAL online
Corresponding Author's E-mail ID: kailashpatidar123@gmail.com

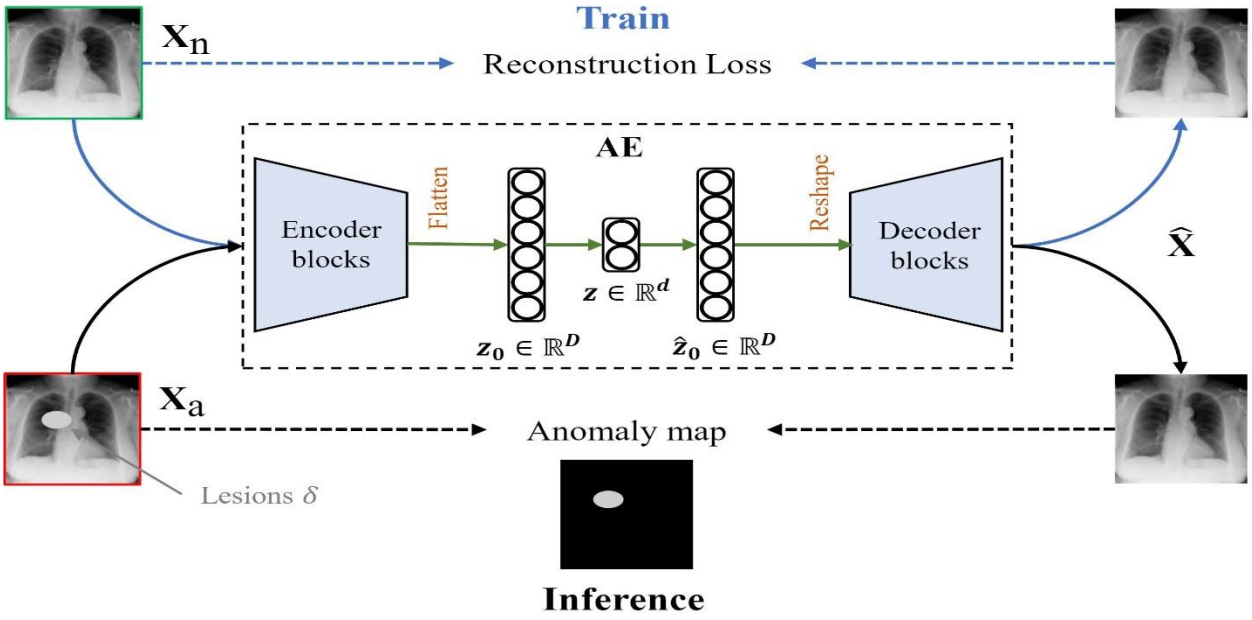

Fig. 1. Conceptual representation of an autoencoder-based anomaly detection framework for cancer data analysis using reconstruction error.

2. Related Work

The evolution of computational techniques for cancer diagnosis has progressed from traditional statistical methods to advanced deep learning models. Early approaches relied on statistical techniques such as logistic regression, linear discriminant analysis, and Bayesian classifiers. These techniques were limited in their ability to handle complicated and high-dimensional biomedical data, although offering fundamental insights. Algorithms including decision trees, support vector machines, and k-nearest neighbors were developed to enhance classification performance as machine learning progressed. These methods demonstrated enhanced capability in capturing nonlinear relationships; however, they remained dependent on feature engineering and balanced datasets. Because deep learning makes it possible to automatically extract features from raw data, it has drastically changed the study of cancer data. Medical imaging and clinical data processing have extensively used neural network architectures, such as convolutional and fully connected networks. Despite their effectiveness, deep learning models frequently need big labeled datasets, which are hard to come by in medical settings. Anomaly detection methods have being investigated more and more in an effort to get around these restrictions. Anomaly detection models, in contrast to traditional classification techniques, concentrate on learning the distribution of normal data and spotting anomalies that can indicate abnormal situations. Reconstruction-based approaches, particularly autoencoders, have shown strong potential in handling high-dimensional and imbalanced datasets. Recent studies demonstrate that autoencoder-based models can effectively capture intrinsic data patterns and identify anomalies through reconstruction error. These approaches offer improved scalability, adaptability, and robustness, making them suitable for cancer data analysis in real-world healthcare environments.

3. Literature Review

3.1 Datasets Used in Cancer Analysis

The effectiveness of deep learning models for cancer detection largely depends on the quality and characteristics of the datasets used. Various publicly available datasets have been widely utilized in research, including medical imaging datasets such as MRI and CT scans, as well as clinical and genomic datasets. Medical imaging datasets are commonly used for tumor detection and classification tasks. These datasets provide high-resolution images that capture structural

and textural information of tissues. To guarantee consistency, they frequently need preprocessing methods including scaling, noise reduction, and normalization. Clinical datasets, including electronic health records, contain structured information such as patient demographics, laboratory results, and medical history. These datasets are useful for predictive modeling but may contain missing values and require data cleaning and imputation techniques. Genomic datasets provide information about gene expression and mutations associated with cancer. These datasets are typically high-dimensional and require dimensionality reduction techniques for effective analysis. A major challenge across all datasets is class imbalance, where abnormal cancer cases are significantly fewer than normal samples. This imbalance necessitates the use of specialized learning techniques such as anomaly detection and data augmentation to improve model performance.

3. 2 Algorithms for Cancer Detection and Analysis

A wide range of machine learning and deep learning algorithms have been applied for cancer detection and analysis. These algorithms vary in how they learn, how much data they need, and how well they can handle complicated biomedical data. For cancer classification tasks, conventional machine learning techniques including Support Vector Machines (SVM), Decision Trees, and k-Nearest Neighbors (k-NN) have been extensively utilized. SVMs have demonstrated excellent performance in binary classification tasks and are especially useful in high-dimensional fields. However, these methods rely heavily on feature engineering and may struggle with complex nonlinear relationships present in medical data. Deep learning models have significantly improved cancer detection by enabling automatic feature extraction. Medical image analysis makes extensive use of Convolutional Neural Networks (CNNs), especially in MRI and CT-based malignancy detection. CNNs can learn spatial hierarchies and extract meaningful features directly from raw images, leading to improved classification performance. Time-series patient records and other sequential medical data have been subjected to Recurrent Neural Networks (RNNs) and their variations. However, compared to CNNs, their use in cancer detection is more restricted. The ability of anomaly detection systems to handle unbalanced datasets has drawn attention in recent years. Autoencoder-based models are particularly useful as they learn compact representations of normal data and detect anomalies through reconstruction error. These models are appropriate for real-world medical applications where anomalous samples are few because they don't require labeled anomaly data. Additional sophisticated methods include generative models like variational autoencoders and adversarial networks, as well as hybrid models that integrate deep learning with conventional machine learning methods. These methods seek to increase the robustness and accuracy of detection in intricate medical datasets.

3.3 Comparative Analysis of Algorithms

Different algorithms used in cancer detection exhibit varying strengths and limitations depending on the nature of the dataset and application requirements. Traditional machine learning models are computationally efficient but depend heavily on feature engineering and may not generalize well for complex data. CNNs and other deep learning models perform better in image-based tasks, but they demand a lot of processing power and big datasets. Autoencoder-based anomaly detection models offer a unique advantage in handling imbalanced datasets by focusing on learning normal data distributions. These models are particularly effective in scenarios where labeled anomaly data is scarce. However, their performance depends on appropriate threshold selection and reconstruction quality.

Table 1. Comparative analysis of commonly used algorithms in cancer detection.

Algorithm	Advantages	Limitations
Support Vector Machine (SVM)	Effective in high-dimensional data; strong classification capability	Requires feature engineering; sensitive to parameter tuning
Decision Trees	Simple and interpretable; low computational cost	Prone to overfitting; lower accuracy for complex data
k-Nearest Neighbors (k-NN)	Easy implementation; no training phase	High computation during testing; sensitive to noise

Convolutional Neural Networks (CNN)	Automatic feature extraction; high accuracy in image analysis	Requires large datasets; computationally expensive
Autoencoder	Effective for anomaly detection; handles imbalanced data well	Depends on reconstruction threshold; less interpretable
Variational Autoencoder (VAE)	Probabilistic modeling; improved generalization	Complex training; computational overhead
Hybrid Models	Combines strengths of multiple approaches	Increased complexity; difficult to optimize

4. Discussion

The comparative analysis of different datasets and algorithms highlights the significant progress made in cancer detection using artificial intelligence techniques. Due to their capacity to automatically learn complicated feature representations, deep learning models—in particular, Convolutional Neural Networks (CNNs) and autoencoder-based architectures—have proven to perform better than conventional machine learning techniques. One of the key observations from the reviewed literature is the strong dependency of model performance on dataset characteristics. High-quality and well-balanced datasets contribute significantly to improved classification accuracy and generalization capability. However, the efficacy of traditional supervised learning techniques is constrained by the inherent imbalance of the majority of real-world medical datasets. In this context, anomaly detection techniques, especially autoencoders, provide a robust solution by focusing on learning normal data distributions rather than relying on labeled abnormal samples. Another important aspect is the role of preprocessing techniques. Methods such as normalization, data augmentation, and feature scaling have been widely used to enhance model performance. These methods aid in lowering data variability and enhancing models' capacity to generalize across various datasets.

Despite these developments, a number of obstacles still exist. Large data sets and powerful computers are frequently needed for deep learning models, which may not be easily accessible in all healthcare settings. Clinical adoption is also significantly hampered by sophisticated models' lack of interpretability. To trust automated diagnostic instruments, medical professionals need systems that are clear and understandable. Future studies should concentrate on creating hybrid models that integrate the advantages of several algorithms, enhancing model interpretability via explainable AI methods, and incorporating multimodal data sources such genomic, clinical, and imaging data. These developments will be essential to improving the dependability and usefulness of AI-based cancer detection systems..

5. Conclusion

This paper presents a comprehensive survey of different datasets and algorithms used in deep learning–based cancer detection and analysis. In order to attain dependable diagnostic performance, the study emphasizes the need of choosing suitable datasets and modeling techniques. Traditional machine learning methods provide baseline solutions but are limited by their dependency on feature engineering and labeled data. On the other hand, deep learning models—in particular, CNNs and autoencoder-based architectures—offer notable benefits for managing intricate and multidimensional biomedical data. Autoencoder-based anomaly detection approaches are especially effective in addressing class imbalance issues by learning normal data distributions and identifying deviations. Additionally, the study highlights how important preprocessing methods and assessment measures are to enhancing model performance. Even though there has been a lot of improvement, issues including data imbalance, computational complexity, and lack of interpretability are still major worries. Deep learning–based approaches have transformed cancer detection and analysis, providing more accurate and efficient diagnostic solutions. To further improve these systems' efficacy and clinical usability, future research should concentrate on explainable AI, hybrid modeling techniques, and the integration of multimodal data.

References

[1] I. T. Sado, “Early multi-cancer detection through deep learning anomaly detection using variational autoencoders,” Journal of Biomedical Informatics, 2024.

[2] H. Zhang et al., “Unsupervised deep anomaly detection for medical images using adversarial autoencoder,” IEEE Medical Imaging Research, 2022.

[3] A. Frotscher et al., “Unsupervised anomaly detection in medical imaging using diffusion models,” Medical Image Analysis, 2025.

[4] P. Sharma et al., “Automated cancer detection using convolutional neural networks: A comprehensive survey,” Neural Networks, 2024.

[5] I. Z. Yao et al., “Deep learning applications in clinical cancer detection,” Journal of Clinical Medicine, 2025.

[6] A. Alloqmani et al., “Deep learning-based anomaly detection framework for breast cancer diagnosis,” Healthcare Analytics, 2023.

[7] S. Park et al., “Unsupervised anomaly detection with generative models for breast cancer screening,” Scientific Reports, 2023.

[8] S. Lu et al., “Patch-wise contrastive learning autoencoder for medical anomaly detection,” Computerized Medical Imaging and Graphics, 2024.

[9] C. Yun et al., “Deep learning anomaly detection in breast ultrasound imaging,” Sensors, 2023.

[10] A. A. Neloy et al., “A comprehensive study of autoencoders for anomaly detection,” Machine Learning with Applications, 2024.

[11] I. Lagogiannis et al., “Deep unsupervised approaches for pathology detection in medical imaging,” arXiv, 2023.

[12] P. Huang et al., “Deep autoencoder-based anomaly detection in radiotherapy treatment plans,” Frontiers in Oncology, 2023.

[13] V. Shukla et al., “Deep learning-based anomaly detection methodologies: A systematic review,” Frontiers in Robotics and AI, 2025.

[14] “Medical anomaly detection using machine learning and deep learning approaches,” International Journal of Intelligent Systems, 2024.

[15] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, 2016.

[16] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, pp. 436–444, 2015.

[17] G. Hinton and R. Salakhutdinov, “Reducing dimensionality with neural networks,” Science, 2006.

[18] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review,” IEEE TPAMI, 2013.

[19] R. Chalapathy and S. Chawla, “Deep learning for anomaly detection: A survey,” ACM Computing Surveys, 2019.

[20] G. Pang et al., “Deep learning for anomaly detection: A review,” ACM Computing Surveys, 2020.

[21] G. Litjens et al., “A survey on deep learning in medical image analysis,” Medical Image Analysis, 2017.

[22] A. Esteva et al., “Dermatologist-level classification of skin cancer using deep neural networks,” Nature, 2017.

[23] A. Hosny et al., “Artificial intelligence in radiology,” Radiology, 2018.

[24] H. E. Kim et al., “AI-assisted cancer detection in mammography,” Radiology, 2019.

[25] Z. Zhang et al., “Deep learning-based cancer detection: A systematic review,” IEEE Access, 2020.