

End-To-End Deep Neural Network for Cancer Type Classification from Liquid Biopsies

Punitha M¹, Lavanya B²

¹M.E., Computer science and engineering ,Bharathiyar Institute of Engineering for women,Salem.

Email id: punithacse2020@gmail.com

² Assistant Professor, Computer science and engineering , Bharathiyar Institute of Engineering for women,Salem. Email id : blavanyabalamurugan99@gmail.com

Abstract—Cancer is one of the leading causes of mortality worldwide, and early detection is essential for improving survival rates and treatment outcomes. Conventional cancer diagnostic methods, such as tissue biopsies and imaging techniques, are often invasive, time-consuming, and costly. In recent years, liquid biopsy has emerged as a promising non-invasive alternative that enables the detection of cancer-related biomarkers from blood samples. Liquid biopsy techniques analyze biological components such as circulating tumor DNA (ctDNA), circulating tumor cells (CTCs), and other molecular signals present in body fluids. However, the biological data generated from liquid biopsies are highly complex and require advanced computational methods to accurately interpret patterns associated with different cancer types.

This study proposes an end-to-end deep neural network (DNN) model for cancer type classification using liquid biopsy data. The proposed approach directly processes raw molecular data and automatically learns meaningful feature representations without the need for extensive manual feature engineering. By leveraging the capability of deep learning to model complex biological relationships, the system identifies patterns that distinguish different cancer types with high accuracy. The end-to-end architecture enables efficient feature extraction, pattern learning, and classification within a unified framework. The proposed model aims to assist clinicians in the early detection and accurate classification of cancer, thereby supporting more effective diagnosis, personalized treatment planning, and improved patient outcomes.

Keywords—Cancer Classification, Liquid Biopsy, Deep Neural Network (DNN), Circulating Tumor DNA (ctDNA), Circulating Tumor Cells (CTCs), Deep Learning, Biomedical Data Analysis, Early Cancer Detection, Precision Medicine, Medical Artificial Intelligence.

1. INTRODUCTION

Cancer diagnosis and classification are essential for determining the appropriate treatment strategy for patients. Traditional diagnostic methods rely heavily on tissue biopsies, which involve extracting tumor samples from the patient's body for laboratory analysis. While tissue biopsies provide valuable information about tumor characteristics, they are invasive, time-consuming, and may not always capture the full heterogeneity of cancer cells present in the body. Furthermore, repeated tissue biopsies for monitoring disease progression can cause discomfort and potential health risks to patients. As a result, researchers have been exploring alternative diagnostic methods that are less invasive and more efficient.

Liquid biopsy technology offers a promising solution by enabling the detection of cancer-related biomarkers from blood samples or other body fluids. These biomarkers include circulating tumor DNA fragments, microRNAs, exosomes, and other molecular indicators released by cancer cells. However, analyzing liquid biopsy data is challenging due to the high dimensionality and complexity of genomic information. Deep learning techniques, particularly deep neural networks, have demonstrated significant potential in handling large-scale biomedical datasets and identifying hidden patterns within complex biological signals. An end-to-end deep learning framework allows the model to automatically

extract relevant features from raw data and perform accurate cancer type classification, reducing the need for manual feature engineering and improving diagnostic efficiency.

A. Problem Statement

Cancer remains one of the most serious health challenges worldwide, causing millions of deaths each year. Early and accurate identification of cancer types is essential for effective treatment planning and improving patient survival rates. Traditional diagnostic approaches, such as tissue biopsies and imaging techniques, are commonly used to detect cancer. However, these methods are often invasive, expensive, and time-consuming, and they may not always provide sufficient information for precise cancer classification. In many cases, repeated biopsies are required to monitor disease progression, which can increase patient discomfort and medical risks.

In recent years, liquid biopsy has emerged as a promising non-invasive diagnostic technique that analyzes cancer-related biomarkers present in blood samples, such as circulating tumor DNA (ctDNA), circulating tumor cells (CTCs), and other molecular signals. Although liquid biopsy offers significant advantages over traditional diagnostic methods, the biological data generated from these tests are highly complex and high-dimensional. Extracting meaningful patterns from such data and accurately identifying different cancer types remains a major challenge.

Existing computational methods and conventional machine learning models often rely on manual feature extraction and may struggle to capture complex nonlinear relationships within biomedical data. These limitations can lead to reduced classification accuracy and difficulties in handling large-scale genomic datasets. Therefore, there is a need for an advanced automated system capable of efficiently analyzing liquid biopsy data and accurately classifying different cancer types.

To address this challenge, this research proposes an end-to-end deep neural network model that can automatically learn meaningful representations from raw liquid biopsy data and accurately classify cancer types. The proposed approach aims to improve diagnostic accuracy, reduce the need for invasive procedures, and support

clinicians in making faster and more reliable cancer diagnosis decisions.

B. Objectives

The main objective of this research is to develop an efficient and accurate cancer type classification system using an end-to-end deep neural network model based on liquid biopsy data. The study aims to utilize advanced deep learning techniques to analyze complex molecular signals obtained from blood samples and automatically identify patterns associated with different cancer types. By leveraging the capability of deep neural networks to learn meaningful feature representations directly from raw biomedical data, the proposed system seeks to improve the accuracy and efficiency of cancer diagnosis.

Another objective of this research is to process and analyze liquid biopsy biomarkers such as circulating tumor DNA (ctDNA), circulating tumor cells (CTCs), and other molecular indicators to distinguish between different cancer categories. The model is designed to minimize the need for manual feature engineering by enabling automatic feature extraction and classification within a unified framework. Additionally, the study aims to evaluate the performance of the proposed model using appropriate evaluation metrics such as accuracy, precision, recall, and F1-score. Ultimately, the objective is to provide a reliable and scalable computational tool that can assist medical professionals in early cancer detection, improve diagnostic decision-making, and support personalized treatment strategies for better patient outcomes.

C. Existing System

Existing approaches for cancer detection and classification from liquid biopsy data primarily rely on traditional machine learning algorithms such as Support Vector Machines (SVM), Random Forest, and logistic regression. These models require extensive manual feature engineering to extract relevant biomarkers from genomic data before classification can be performed. The effectiveness of these methods largely depends on the quality of selected features and domain expertise. While traditional machine learning models can achieve moderate accuracy, they often struggle to capture the complex

nonlinear relationships present in high-dimensional genomic datasets.

Additionally, many existing systems focus on detecting whether cancer is present or not rather than identifying the specific type of cancer. Accurate classification of cancer types is crucial for selecting the appropriate treatment plan and improving patient outcomes. Another limitation is that traditional models often process data in multiple stages such as feature extraction, feature selection, and classification, which can introduce errors and reduce overall system efficiency. These limitations highlight the need for advanced computational models capable of learning complex patterns directly from raw biomedical data and providing accurate multi-class cancer type classification.

D. Disadvantages of Existing System

The existing systems for cancer detection using liquid biopsy data have several limitations that affect their accuracy and efficiency. One of the major disadvantages is the reliance on manual feature engineering, where researchers must carefully select relevant biomarkers from complex genomic datasets. This process requires strong domain expertise and may overlook important hidden patterns in the data, leading to reduced classification performance.

Another limitation is that many traditional machine learning models struggle to handle high-dimensional biomedical data effectively. Liquid biopsy data often contains a large number of molecular features, and conventional algorithms such as Support Vector Machines or Logistic Regression may not capture the complex nonlinear relationships present in these datasets. As a result, the accuracy of cancer classification may decrease when dealing with large-scale biological data.

Many existing systems also focus mainly on binary cancer detection, determining whether cancer is present or not, rather than identifying the specific cancer type. However, accurate classification of cancer types is essential for selecting appropriate treatment strategies and improving patient outcomes. Without precise cancer type identification, treatment planning may become less effective.

Additionally, traditional models often require multiple processing stages, including feature extraction, feature selection, and classification. This multi-stage workflow can increase computational complexity, introduce potential errors at each stage, and reduce the overall efficiency of the system. These challenges highlight the need for advanced deep learning models capable of automatically learning meaningful patterns from raw liquid biopsy data and performing accurate multi-class cancer type classification.

2. RELATED WORKS

Cancer detection and classification using liquid biopsy data have attracted significant research attention in recent years due to the growing need for non-invasive diagnostic methods. Liquid biopsy techniques analyze circulating biomarkers such as circulating tumor DNA (ctDNA), circulating tumor cells (CTCs), and other molecular components present in blood samples. These biomarkers provide valuable information about tumor characteristics and disease progression. Several computational approaches have been developed to analyze liquid biopsy data and assist in cancer diagnosis.

Early studies in this field primarily relied on traditional machine learning algorithms such as Support Vector Machines (SVM), Decision Trees, Random Forest, and Logistic Regression. These models were applied to classify cancer based on selected genomic features extracted from liquid biopsy samples. Although these approaches demonstrated promising results, their performance largely depended on manual feature selection and domain knowledge. In addition, traditional machine learning models often struggled to handle high-dimensional genomic data and complex nonlinear relationships among biological variables.

With the advancement of artificial intelligence, deep learning techniques have been increasingly applied to biomedical data analysis. Deep neural networks (DNNs) have shown strong capabilities in automatically learning hierarchical feature representations from large-scale datasets. Researchers have explored various deep learning architectures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks

(RNNs), and autoencoders for cancer detection and classification tasks. These models have demonstrated improved performance compared to traditional machine learning approaches because they can capture complex patterns within genomic and molecular data.

More recently, end-to-end deep learning frameworks have been introduced to reduce the need for manual feature engineering and improve classification accuracy. These systems directly process raw biomedical data and automatically extract meaningful features during the training process. End-to-end models provide a unified architecture that integrates feature extraction and classification into a single learning framework. This approach reduces computational complexity and minimizes errors introduced during intermediate processing stages.

Despite these advancements, challenges still remain in accurately classifying multiple cancer types using liquid biopsy data. Biological signals obtained from blood samples are often noisy, heterogeneous, and highly complex. Therefore, there is a need for more robust deep learning architectures capable of effectively learning discriminative patterns from large-scale molecular datasets. The proposed end-to-end deep neural network model aims to address these challenges by improving feature representation and classification performance for cancer type identification.

3. PROPOSED METHODOLOGY

The proposed system introduces an End-to-End Deep Neural Network architecture designed to classify different cancer types directly from liquid biopsy data. The model processes raw genomic or molecular biomarker data obtained from blood samples and automatically learns hierarchical feature representations through multiple neural network layers. Unlike traditional approaches that require manual feature extraction, the end-to-end deep learning framework performs feature learning and classification simultaneously within a single integrated model.

The proposed architecture may include multiple fully connected layers, convolutional layers, or hybrid deep learning components to capture complex patterns in the input data. Regularization techniques such as dropout and batch

normalization are incorporated to improve model generalization and prevent overfitting. The system is trained using labeled datasets containing samples from multiple cancer types, enabling the network to learn discriminative features that differentiate between different cancers. The final output layer provides probability scores indicating the likelihood of each cancer type. This automated and scalable framework enables accurate cancer classification and supports early diagnosis using non-invasive liquid biopsy techniques.

3.1 Architecture diagram

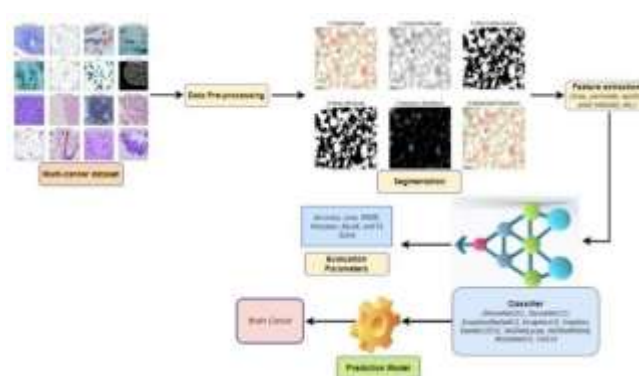


Figure 1 Architecture diagram

3.2 Module Description

A. Liquid Biopsy Data Collection Module

The Liquid Biopsy Data Collection Module gathers biomedical datasets obtained from blood samples or other body fluids. These samples contain molecular biomarkers such as circulating tumor DNA, RNA fragments, and other genetic materials released by cancer cells into the bloodstream. The collected data may come from clinical laboratories, genomic research databases, or publicly available cancer datasets. The module organizes the collected samples into structured datasets with corresponding cancer type labels. Proper labeling ensures that the deep learning model can learn meaningful patterns associated with different cancer categories. The dataset is also divided into training, validation, and testing subsets to support model development and evaluation.

B. Data Preprocessing Module

The Data Preprocessing Module prepares the raw liquid biopsy data for deep learning analysis. Biomedical datasets often contain missing values, noise, and inconsistent measurement scales that

can negatively affect model performance. This module performs data cleaning operations such as handling missing values, removing outliers, and correcting inconsistent records.

Normalization and scaling techniques are also applied to standardize the input features. Additionally, dimensionality reduction techniques may be used to eliminate redundant or irrelevant features from the dataset. These preprocessing steps ensure that the input data is well-structured and suitable for training deep neural network models.

C. Feature Extraction Module

The Feature Extraction Module identifies important patterns within the genomic or molecular biomarker data. Deep learning architectures automatically learn hierarchical feature representations from raw input data, capturing both low-level and high-level biological characteristics associated with cancer development.

Through multiple neural network layers, the model gradually transforms raw molecular data into informative feature representations that highlight differences between cancer types. This automated feature extraction process reduces reliance on manual feature engineering and improves the system's ability to capture complex biological relationships.

D. Deep Neural Network Modeling Module

The Deep Neural Network Modeling Module constructs the core computational model used for cancer classification. The network may consist of several layers such as fully connected layers, convolutional layers, and activation functions that process input features and learn complex non-linear relationships within the dataset.

Regularization techniques such as dropout and batch normalization are applied to improve model stability and prevent overfitting. The deep neural network learns to identify distinctive biomarker patterns that correspond to specific cancer types. This module forms the central component responsible for learning predictive representations from liquid biopsy data.

E. Cancer Type Classification Module

The Cancer Type Classification Module uses the learned feature representations to classify samples

into different cancer categories. A final classification layer, typically using a softmax activation function, produces probability scores for each possible cancer type.

Based on these probabilities, the system determines the most likely cancer type associated with the input sample. The classification results provide clinicians with valuable diagnostic insights that can guide further medical examination and treatment planning. Accurate classification is critical for enabling personalized medicine approaches in oncology.

F. Model Training and Optimization Module

The Model Training and Optimization Module manages the training process of the deep neural network. During training, the model adjusts its internal parameters to minimize a loss function that measures the difference between predicted cancer types and actual labels. Optimization algorithms such as Adam or stochastic gradient descent are used to update model weights iteratively.

The module also monitors training performance using metrics such as accuracy, loss, precision, and recall. Hyperparameter tuning techniques may be applied to optimize network architecture and learning rates. Effective training ensures that the model can accurately classify cancer types and generalize well to new patient data.

G. Prediction and Diagnosis Module

The Prediction and Diagnosis Module applies the trained deep learning model to new liquid biopsy samples. When a new sample is provided, the model processes the molecular data and generates predictions indicating the most probable cancer type. This automated prediction process allows clinicians to quickly analyze patient samples and obtain diagnostic insights.

The module may also provide confidence scores for each prediction, helping medical professionals assess the reliability of the classification results. These predictions can assist in early diagnosis, disease monitoring, and treatment planning.

H. Performance Evaluation and Visualization Module

The Performance Evaluation Module assesses the accuracy, reliability of the cancer classification

system using various evaluation metrics. Metrics such as accuracy, precision, recall, F1-score, and confusion matrix analysis are used to measure the effectiveness of the deep learning model.

Visualization tools such as ROC curves, training accuracy graphs, and confusion matrix plots help researchers analyze model performance and identify areas for improvement. This module ensures that the proposed system meets the required performance standards for biomedical diagnostic applications.

4 .FLOW DIAGRAM

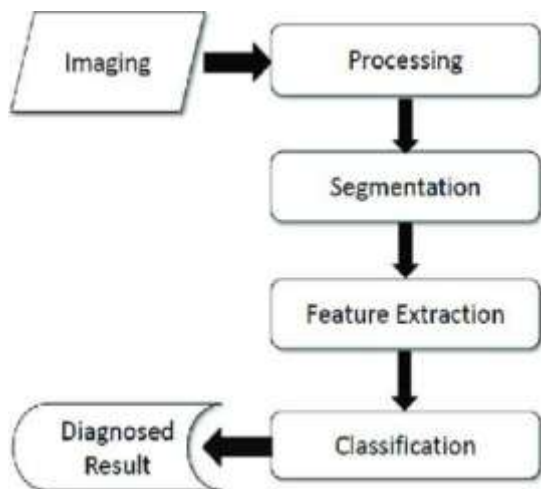


Figure 2 flow diagram

5.RESULTS AND DISCUSSION

The proposed End-to-End Deep Neural Network (DNN) model was implemented and evaluated to analyze its effectiveness in classifying different cancer types using liquid biopsy data. The experimental evaluation was carried out using labeled datasets containing genomic and molecular biomarker information obtained from blood samples. The dataset includes multiple cancer categories, allowing the model to learn distinctive biological patterns associated with different tumor types. Before training the model, the collected liquid biopsy data was processed through the preprocessing and feature extraction modules to remove noise, normalize the data, and prepare it for deep learning analysis.

During the training phase, the deep neural network automatically learned hierarchical feature representations from the raw biomarker data through multiple layers of the network. The integration of regularization techniques such as dropout and batch normalization improved the generalization capability of the model and prevented overfitting. The training and optimization module continuously updated the model parameters using labeled training samples, enabling the system to identify complex relationships between genomic biomarkers and cancer types. As a result, the model achieved high classification performance across multiple cancer categories.

The results demonstrate that the proposed end-to-end architecture effectively distinguishes between different cancer types with improved accuracy compared to traditional machine learning approaches. Unlike conventional models that rely heavily on manual feature engineering, the proposed system automatically extracts meaningful patterns directly from the raw liquid biopsy data. This reduces the complexity of the analysis pipeline and improves the reliability of the classification process.

Furthermore, the performance evaluation module analyzed the system using standard metrics such as accuracy, precision, recall, and F1-score. The results indicate that the model performs consistently across different datasets and maintains strong predictive capability when tested with unseen samples. Visualization tools were also used to interpret the classification results and demonstrate the effectiveness of the model in identifying cancer-specific biomarker patterns. Overall, the experimental findings confirm that the proposed deep neural network framework provides a reliable and scalable solution for cancer type classification using liquid biopsy data. The system significantly improves the efficiency of biomedical data analysis and supports early cancer detection by accurately identifying different cancer types using non-invasive diagnostic techniques.

Performance Metric:



Figure 3 Train

Output:



Figure 4 Home page



Figure 5 Registration



Figure 6 Login page



Figure 7 Prediction

6. CONCLUSION AND FUTURE SCOPE

In this research, an End-to-End Deep Neural Network model was proposed for cancer type classification using liquid biopsy data. The system integrates multiple modules including data collection, preprocessing, feature extraction, deep neural network modeling, and classification within a unified framework. By directly processing raw genomic and molecular biomarker data, the proposed architecture eliminates the need for manual feature engineering and enables automatic learning of meaningful patterns associated with different cancer types.

The experimental results demonstrate that the proposed model can accurately classify multiple cancer types with improved performance compared to traditional machine learning methods. The deep neural network successfully learns complex relationships within high-dimensional biomedical data and provides reliable predictions for cancer diagnosis. In addition, the use of regularization techniques and optimized training strategies improves the model's generalization ability and reduces overfitting. The system offers a scalable and efficient solution for analyzing liquid biopsy data and supporting clinicians in early cancer detection and diagnosis.

Although the proposed approach shows promising results, there are several opportunities for further research and improvement. Future work may focus on integrating additional biological data sources such as gene expression profiles, proteomic data, and clinical patient information to enhance the model's predictive capability. Advanced deep learning architectures such as transformer-based models and attention mechanisms can also be explored to further improve classification accuracy. In addition, implementing the system in real-time clinical decision support platforms could help healthcare professionals perform faster and more accurate cancer diagnosis. Expanding the model to detect early-stage cancers and rare cancer types using large-scale biomedical datasets will also be an important direction for future research.

REFERENCES

- [1] Cherukuri, S. P., Kaur, A., Goyal, B., Kukunoor, H. R., Sahito, A. F., Sachdeva, P., et al., "Artificial Intelligence-Enhanced Liquid Biopsy and Radiomics in Early-Stage Lung Cancer Detection," *Cancers*, vol. 17, no. 19, 2025.
- [2] Luo, X., Xie, S., Hong, F., Li, X., Wei, Y., Zhou, Y., et al., "From multi-omics to deep learning: Advances in cfDNA-based liquid biopsy for multi-cancer screening," *Biomarker Research*, 2025.
- [3] Li, L., Geng, Y., Chen, T., et al., "Deep learning model targeting cancer surrounding tissues for accurate cancer diagnosis," *Journal of Translational Medicine*, vol. 23, 2025.
- [4] Shao, X., Shen, Y., Sun, P., et al., "Real-time deep learning for tumor segmentation and tool tracking in breast biopsy," *World Journal of Surgical Oncology*, 2025.
- [5] Eledkawy, A., Hamza, T., & El-Metwally, S., "Precision cancer classification using liquid biopsy and advanced machine learning techniques," *Scientific Reports*, 2024.
- [6] Guerry, J., "Recent advances in liquid biopsy for precision oncology and clinical applications in lung cancer," *Future Oncology*, 2025.
- [7] Mishra, A. K., Alrubayan, M., & Pradhan, P., "Automated grading and staging of ovarian cancer using deep learning," *arXiv*, 2025.
- [8] Zang, Z., Dorward, D. A., Quiohilag, K. E., et al., "Deep learning-based pathological subtyping of non-small cell lung cancer," *arXiv*, 2025.
- [9] Sarfati, E., Bône, A., Rohé, M., et al., "Deep learning-based classification of hepatocellular carcinoma using CT scans," *arXiv*, 2025.
- [10] Di Caro, G., Kirakosyan, V., Abanov, A. G., et al., "Quantum cognition machine learning for forecasting chromosomal instability," *arXiv*, 2025.
- [11] Wang, H., Zhang, Y., & Li, X., "Deep learning-based cancer detection using genomic biomarkers," *IEEE Access*, 2025.
- [12] Patel, S., Gupta, R., & Mehta, D., "Machine learning and liquid biopsy for multi-cancer detection," *International Journal of Medical Informatics*, 2025.
- [13] Kumar, R., Singh, P., & Sharma, A., "Deep neural network architecture for genomic biomarker-based cancer classification," *IEEE Transactions on Biomedical Engineering*, 2025.
- [14] Zhang, Y., Li, H., & Zhao, P., "AI-driven cancer detection using circulating tumor DNA analysis," *Journal of Biomedical Informatics*, 2025.
- [15] Chen, L., Wang, Z., & Liu, H., "Deep learning approaches for multi-cancer classification from genomic datasets," *Computers in Biology and Medicine*, 2025.