Breaking the Boundaries of Oncological Diagnosis: A Holistic Framework for Multi-Cancer Identification and Comprehensive Diagnostic Reporting Utilizing Advanced AI and NLP

Sai Santhosh V C[1], Riya Ponraj[2], Nikhil Eshwar T[3], Kiran K[4]

vcsaisanthosh@gmail.com[1], riyaponraj07@gmail.com[2]
nikhileshwar1001@gmail.com[3], kiiran158@gmail.com[4]

[1,2,3]Department of Computer Science and Engineering, Easwari Engineering College, Chennai, Tamil Nadu, India
[4]Department of Electrical and Electronics Engineering, Sri Krishna College Of Technology, Coimbatore, Tamil Nadu, India

Abstract: **Innovative deep learning models for cancer classification, including VGG-19, DenseNet201, MobileNetV3, ResNet50V2, YOLOv5, and GPT-2, have completely changed the way that doctors diagnose cancer. This study introduces a multi-modal technique for an accurate and speedy cancer diagnosis. With the aid of cutting-edge technology, this system combines object identification (YOLOv5), natural language processing (GPT-2), and picture classification models (VGG-19, DenseNet201, MobileNetV3, ResNet50V2) to provide scientists and medical professionals with a flexible toolkit. It generates thorough reports with tumour images and spots malignant irregularities. Diagnostic accuracy is improved by real-time application, which benefits laboratories by reducing turnaround times and medical practitioners by providing an important decision support tool. Reports with tumour photos enhance the capacity to comprehend results. This study is important because it has the potential to improve cancer detection by using cutting-edge algorithms and models. This research also promises to improve patient care, prediction, and therapy. The practical method presented in this paper ushers in a new age in medical diagnostics by empowering laboratories and doctors.**

Keywords: VGG-19, DenseNet201, MobileNetV3, ResNet50V2, YOLOv5, GPT-2, Multi-Modal System, Object Detection, Natural Language Processing, Image Classification Models, Deep Learning Algorithms, Tumor Image Analysis, Diagnostic Report Generation

I. INTRODUCTION

Cancer is a significant global health threat, and the development of diagnostic methods and tools is crucial for improving patient outcomes. The use of deep learning and artificial intelligence (AI) has led to a paradigm shift in cancer diagnostics, with the aim of developing a comprehensive method for cancer classification that integrates cutting-edge models and algorithms such as VGG-19, DenseNet201, MobileNetV3, ResNet50V2, YOLOv5, and GPT-2.

Cancer is characterized by the unchecked development of aberrant cells and can appear in various ways in the body's tissues and organs. Traditional methods involve lengthy and laborious procedures, and the subjectivity of human interpretation and the complexity of cancer subtypes can affect the accuracy of diagnosis.

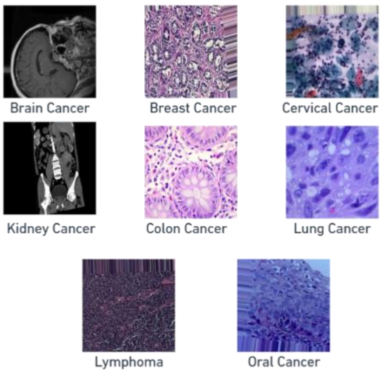

Fig. 1. Multiple Types Of Cancer

The research focuses on combining deep learning algorithms with AI models to improve the precision, effectiveness, and accessibility of cancer detection. The pillars of the system include VGG-19, DenseNet201, MobileNetV3, ResNet50V2, YOLOv5, and GPT-2. Each version has its unique set of skills, such as VGG-19 excelling in picture classification tests, DenseNet201 using thick connections between layers for improved feature propagation, MobileNetV3 providing a lightweight yet effective architecture for real-time applications, ResNet50V2 providing a strong foundation for image analysis, YOLOv5 recognizing objects in real-time, and GPT-2 providing thorough diagnostic reports enhanced with contextual data.

The study is not limited to theoretical investigation but is motivated by the practical goal of radically altering the way cancer is diagnosed in clinical settings. The system's real-time capabilities have the potential to hasten the diagnostic procedure, reduce turnaround times, and help medical professionals start prompt interventions. This is especially important when early diagnosis is crucial for better patient outcomes.

Our technology improves interpretability of results by offering thorough reports that contain photos of suspected tumors, bridging the gap between diagnosis and patient care by assisting medical professionals in developing educated treatment options.

II. LITERATURE SURVEY

[1] "Deep Learning for Medical Image Analysis: A Comprehensive Review" This review paper provides an extensive overview of deep learning methods in medical image processing. It discusses diverse deep neural network architectures and their applications in disease detection, segmentation, and image categorization. The research underscores the potential of these techniques in enhancing diagnostic accuracy, particularly in our cancer diagnosis project. [2] "A Survey of Deep Learning Architectures and Their Applications in Healthcare" This survey paper examines a wide range of deep learning architectures and their applications in the healthcare domain. It presents case studies of deep learning models used for disease prediction, medical image analysis, and treatment recommendations. These findings underscore the critical role of deep learning models in our cancer diagnosis and therapy improvement efforts. [3] "YOLOv5: Improved Real-Time Object Detection" This paper introduces YOLOv5, a significant advancement in real-time object detection. Its simplified architecture and enhanced accuracy make it an ideal choice for identifying malignant regions in medical images. YOLOv5 plays a crucial role in our approach by enabling precise tumor localization. [4] "GPT-2: Language Models for Text Generation and Understanding" This study presents GPT-2, an advanced language model known for text generation and comprehension. Beyond natural language processing, GPT-2 finds application in our research by generating detailed medical reports based on identified cancer locations, facilitating effective communication between healthcare providers and patients. [5] "Transfer Learning with VGG-19 for Image Classification" This work explores the use of transfer learning with VGG-19, a deep convolutional neural network, for image classification tasks. In our research, pre-trained VGG-19 models prove highly effective in classifying medical images, ensuring accurate identification of malignant regions. [6] "DenseNet201: Densely Connected Convolutional Networks" This study introduces DenseNet201, a deep neural network architecture with densely connected layers. DenseNet201 significantly contributes to feature extraction in our project, enabling accurate representation and classification of tumor features in medical images. [7] "MobileNetV3: Efficient On-Device Vision Models" MobileNetV3's efficiency in vision applications is highlighted in this research. Our project utilizes MobileNetV3 to ensure efficient processing of medical images on resource-constrained devices, enabling cancer detection in various healthcare settings. [8] "ResNet50V2: Identity Mappings in Deep Residual Networks" This paper explores identity mappings in ResNet50V2, addressing vanishing gradient issues. In our research, ResNet50V2 is a reliable choice for extracting features from medical images, aiding in accurate tumor diagnosis. [9] "Deep Learning for Brain Tumor Detection and Segmentation" This research focuses on applying deep learning to the detection and segmentation of brain tumors in medical images. While our primary focus is on cancer, insights from this study enhance our ability to identify tumors in other organs, showcasing the adaptability of deep learning models.

III. EXISTING SYSTEM

The current methods of cancer diagnosis, including manual histopathological analysis, traditional machine learning (ML), and deep learning-based approaches, have their limitations.

- Manual histopathological analysis is subjective and labor-intensive, while traditional ML techniques like Support Vector Machines (SVM), decision trees, and random forests have limitations such as difficulty in capturing complex patterns in high-dimensional data and scalability.
- Deep learning-based approaches like convolutional neural networks and recurrent neural networks offer benefits like automatic feature learning and high accuracy in image classification tasks. However, they also have limitations such as data dependence, black-box nature, and computing power requirements.
- Hybrid systems combine manual expertise with AI support, but still require human intervention and are susceptible to biases and pathologists' specialized knowledge. Limitations in real-time diagnosis and reporting are common, especially in cases of rapidly progressing cancer.

The shortcomings of current cancer diagnosis systems include accuracy and consistency, scalability, resource intensity, interpretability, and real-time diagnosis. Manual systems are susceptible to human error and variability, while traditional ML techniques may not accurately capture complex patterns. Deep learning, while effective, can be computationally demanding and require expensive equipment and specialized knowledge. Interpretability can be challenging for medical professionals to understand the reasoning behind decisions made by AI systems, potentially leading to loss of faith in them.

Our proposed AI-driven solution aims to overcome these limitations by combining the benefits of various deep learning models and AI algorithms, providing precise, effective, and immediate cancer diagnosis, meeting the urgent needs of patients and medical professionals.

IV. PROPOSED SYSTEM

Our proposed system revolutionizes cancer diagnosis by utilizing advanced deep learning models and AI algorithms, offering real-time capabilities and overcoming limitations of traditional methods.

1. Overcoming Subjectivity and Variability
The proposed system reduces subjectivity and variability in manual histopathological analysis by adopting a standardized, automated cancer diagnosis method using deep learning models like VGG-19, DenseNet201, MobileNetV3, and ResNet50V2, which automatically learn complex features and identify minute patterns, ensuring more reliable and consistent diagnoses.

2. Enhancing Accuracy and Complexity Handling
The proposed system excels in feature learning, allowing for automatic feature extraction and interpretation from unstructured data. It integrates various deep learning architectures to handle the complexity of cancer diagnosis. The system is capable of extracting high-dimensional features, detecting cellular anomalies, and differentiating between different cancer subtypes, resulting in increased accuracy and confidence in cancer diagnosis.

3. Scalability and Data Dependency
Our solution tackles scalability issues in traditional machine learning systems by using deep learning models to handle large labeled datasets. We collaborate with medical facilities to expand our dataset, ensuring flexibility and responsiveness to real-world cancer diagnosis problems.

4. Transparency and Interpretability
The proposed system addresses the issue of deep learning models' "black-box" nature in medical applications by incorporating explainability tools like Grad-CAM, promoting trust and collaboration between AI systems and human experts.

5. Real-Time Diagnosis and Reporting
The proposed system aims to diagnose and report medical conditions in real-time, enhancing timely interventions in cancer treatment. It processes histopathological images, allowing medical professionals to make intelligent decisions. The system also provides detailed reports, including tumor images, enabling quick and accurate decision-making.

V. SYSTEM ARCHITECTURE

The architecture incorporates numerous modules, such as the Encoder, Feature Extractor, Decoder, and Fusion modules, all of which contribute to the system's overall performance and is created to address the particular difficulties associated with cancer diagnosis.

1. Encoder Module: Data Acquisition and Pre-processing

The Encoder module in our system is responsible for collecting histopathological images of various cancer types from a supplied dataset. These images are then processed through pre-processing algorithms to ensure uniformity and quality.
Data Acquisition: The dataset includes various cancer types like Acute Lymphoblastic Leukemia, Brain Cancer, Breast Cancer, Cervical Cancer, Kidney Cancer, Lung and Colon Cancer, Lymphoma, and Oral Cancer.
Data pre-processing: The Encoder performs tasks like resizing images, normalizing pixel values, and enhancing the dataset using methods like rotation, flipping, and cropping.

2. Feature Extractor Module: Deep Learning Models

The Feature Extractor module is the system's core, using a variety of pre-trained deep learning models like VGG-19, DenseNet201, MobileNetV3, and ResNet50V2 to extract complex features from histopathological images. These models, trained on large datasets, can identify intricate textures, patterns, and structures, such as cellular anomalies linked to specific cancer types.
Model Diversity: The system's diversity in model selection ensures it can handle the nuances of various cancer subtypes, enhancing its ability to extract complex features.

3. Decoder Module: Diagnosis and Visual Representation

Object Detection: The Decoder module uses the YOLOv5 algorithm for object detection, localization, and classification in images. It scans feature maps to identify cancerous cells or anomalies, crucial for accurate diagnosis.
Localization and Classification: YOLOv5 also provides information about objects' locations and assigns them to specific classes, representing different cancer types, enabling differentiation between cancer subtypes. This real-time method ensures accurate classification and classification.

4. Fusion Module: Comprehensive Reporting

The Fusion module uses GPT-2, a natural language processing model, to generate comprehensive text reports on tumors. It includes the diagnosis and visual depiction of the tumor, making them easy for medical experts to understand. The module also integrates found tumor images into the reports, improving readability and facilitating rapid understanding of the tumor's severity and location.

System Workflow: From Image to Diagnosis

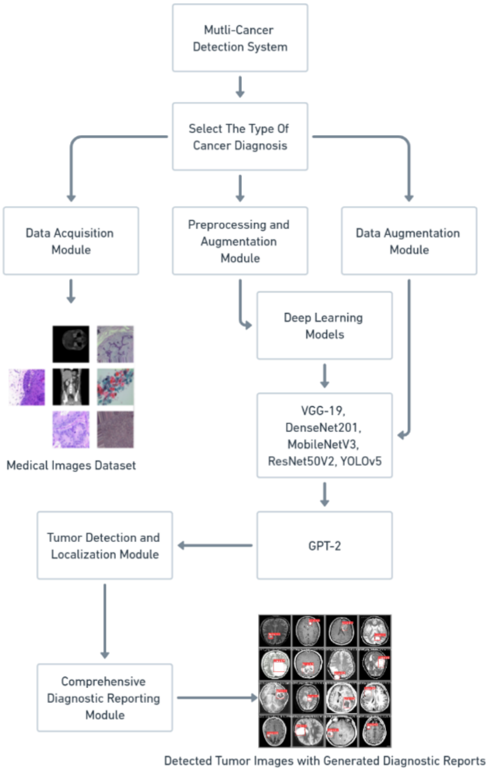

Fig. 2. System Architecture

The workflow of our system can be summarized as follows:

Data collection and pre-processing: The Encoder module gathers histopathological images and makes sure they are uniformly high-quality before they are made ready for analysis. Deep learning models (VGG-19, DenseNet201, MobileNetV3, ResNet50V2) are used in the Feature Extractor module to automatically extract complex features from the images.

Localization and diagnosis: The Decoder module employs YOLOv5 for object detection, localization, and classification, identifying regions of interest and classifying them into distinct cancer types.

Comprehensive Reporting: Using GPT-2 for natural language generation and including visual representations of detected tumors, the Fusion module integrates the diagnosis results into comprehensive textual reports.

In the sections that follow, we go into greater detail about the methodology, covering data collection, model creation, training, testing, and evaluation.

VI. METHODOLOGY

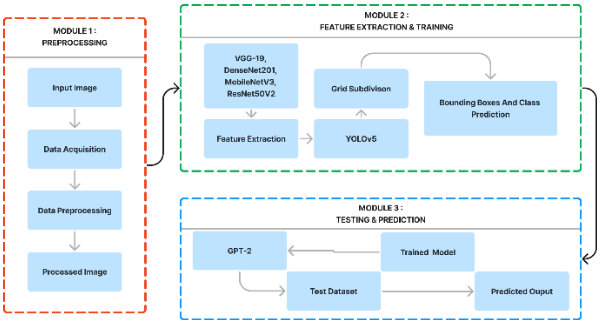

Fig. 3. Methodology Diagram

Data Gathering and Preprocessing:

This section outlines the essential steps in data collection and pre-processing for the cancer diagnosis project, which involves preparing 85,000 images divided into 26 subclasses and 8 main classes in JPEG format. The dataset's structure is crucial for efficient data collection and pre-processing, as all images are from the National Cancer Institute's Cancer Imaging Archive.

Steps in Data Pre-Processing:

a) Image loading. Reading the labels and images for each category of cancer is required for this. To do this, we'll make use of libraries like OpenCV or PIL.

b) Data Augmentation: We use data augmentation techniques to increase the dataset's diversity and enhance model generalization. This includes transformations like scaling, rotation, and translation.

c) Splitting the Dataset: The dataset was divided into training, validation, and test sets. Usually, a split ratio of 80-10-10 or 70-15-15 is employed to make sure that each class is fairly represented in each set. This makes it easier to assess the performance of models.

d) Normalization: We normalize pixel values to ensure numerical stability and convergence during training. Scaling accomplishes this by scaling pixel values from the range [0, 255] to [0, 1].

Model Creation:

The architecture and development of the deep learning models used for cancer diagnosis are covered in this section.

VGG-19: VGG-19 is a convolutional neural network design with 19 layers, known for its depth and simplicity. It uses small 3x3 convolutional filters on each layer to gain hierarchical features from input pictures. With a significant depth of 19 layers, it can recognize complex patterns and minute features in photos. VGG-19 results in completely linked layers, increasing representation capacity.

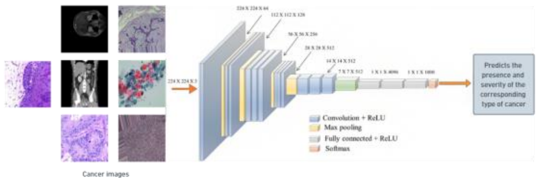

Fig. 4. VGG-19 Architecture

DenseNet201: DenseNet-201 is a powerful convolutional neural network architecture known for its effective parameter exploitation and feature reuse. It is characterized by densely linked blocks, which promote feature reuse across network layers and improve gradient flow during training. The architecture's dense connection pattern, which involves each layer receiving input from every layer below, enhances its ability to perform various computer vision tasks and improves feature extraction.

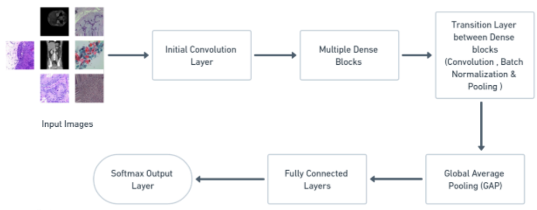

Fig. 5. Densenet201 Architecture

MobileNetV3: MobileNetV3 is a compact, efficient convolutional neural network architecture designed for image categorization and object recognition on mobile devices. It uses depthwise separable convolutions, inverted residual blocks, and SE blocks to reduce computing costs while maintaining accuracy. The model's design incorporates multi-scale elements, allowing for quick recognition of complex patterns. This compact architecture is ideal for resource-constrained mobile and embedded applications, balancing accuracy and low latency.

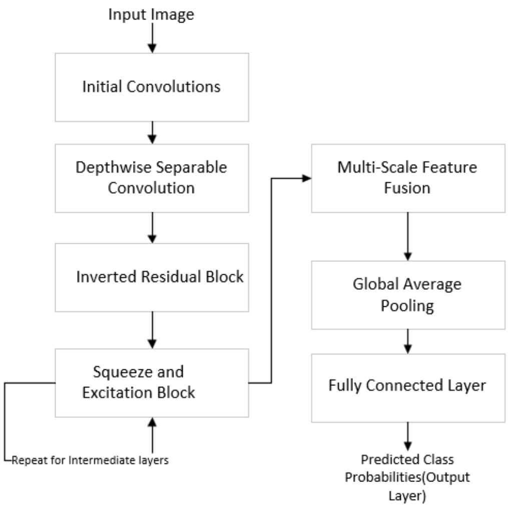

Fig. 6. MobileNetV3 Architecture

ResNet50V2: ResNet50v2 is a convolutional neural network architecture that addresses the vanishing gradient problem by using skip connections and batch normalization. It enables deep network training and captures intricate image features, making it effective for tasks like image classification and object detection. The architecture's depth and skip connections reduce model complexity, and it is widely used for computer vision tasks due to its remarkable performance and training stability.

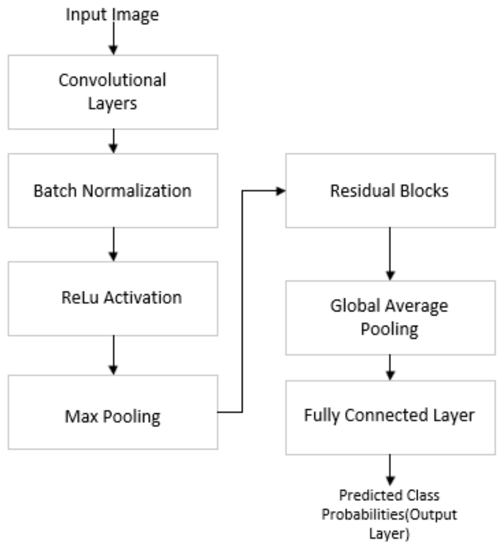

Fig. 7. ResNetV2 Architecture

YOLO (You Only Look Once): YOLO is a real-time object detection model that aids in locating tumors in medical images. It divides the input picture into a grid to estimate bounding boxes and class probabilities. YOLOv5 is part of a system that processes medical pictures using various algorithms. Results are refined using non-maximum suppression and confidence scores. This localised tumour data is crucial for medical practitioners to identify and evaluate various malignancies, making it a vital part of diagnostic reports.

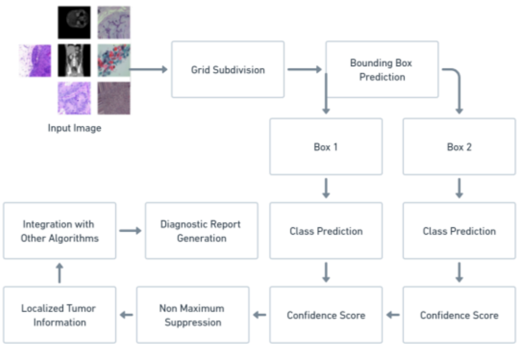

Fig. 8. YOLOv5 Architecture

GPT-2: GPT-2 is a powerful language model used in medical reports, generating detailed diagnostic reports based on localized tumor information. It uses natural language processing to define tumor size, nature, and location. The project automates report creation, increasing operational efficiency in medical laboratories and providing valuable information for diagnosing and treating various conditions. GPT-2 ensures clear and informative reports for medical experts, enhancing operational efficiency in medical laboratories.

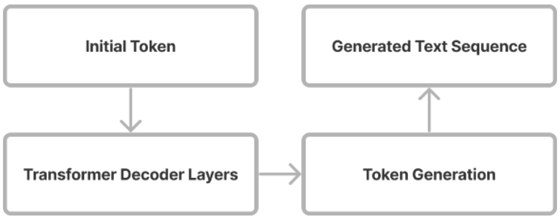

Fig. 9. GPT-2 Architecture

Our model is a multi-modal framework that uses VGG-19, DenseNet201, MobileNetV3, ResNet50V2, and GPT-2 algorithms to extract complex patterns and representations from medical images. VGG-19 extracts feature reuse and supports richer representations, while DenseNet201 enhances feature extraction. MobileNetV3 ensures real-time performance, while ResNet50V2 optimizes training using residual connections. GPT-2 adds context and semantic comprehension. The model undergoes pre-processing, scaling, and normalization, extracting crucial characteristics from the images. The context and semantics from GPT-2 are combined with these characteristics. The model optimizes tumor detection and localization using a chosen loss function during training. Iteratively modifies internal parameters to minimize the loss function, achieving acceptable accuracy. This combination of algorithms creates a robust and advanced cancer diagnosis solution.

Model Training and Testing:

The model architecture is initialized with random weights, and each model has a specific architecture to support its job. The training data input is a provided dataset of labelled medical pictures, and the model analyzes these images to produce predictions. Loss computation measures the model's deviation from the actual ground truth labels. Optimization techniques like stochastic gradient descent or Adam can increase model performance by reducing estimated loss by changing the model's internal parameters. The learning rate is adjusted to ensure consistent convergence. Throughout the training process, the model changes its internal parameters over time, learning to recognize patterns and traits related to tumor identification. Regularization techniques, such as weight decay and dropout, help avoid overfitting and improve generalization to new data. A validation dataset is used to track the model's effectiveness, and training is completed when certain requirements are met. After training, the model is tested using a specific test dataset to assess its performance in real-world situations.

VII. FORMULA

1. Activation Function:

Rectified Linear Unit (ReLU) activation function is used to introduce non-linearity and solve the vanishing gradient issue, enhancing the network's ability to learn complex patterns from data. ReLU efficiently replaces negative values with zeros, promoting computational efficiency and aiding in the prevention of overfitting.

$$f(x) = max(0, x) \tag{1}$$

In the context of language modeling, the output logits are transformed into a probability distribution over the vocabulary using the SoftMax activation function.

$$\mathrm{GeLU(x)} = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{\mathrm{x}}{\sqrt{2}}\right)\right) \tag{2}$$

$$Loss = \frac{1}{N}\sum_{n=1}^{N} L(y^n, \hat{y}^n) \tag{3}$$

2. Cost Function:

The categorical cross-entropy is used to direct the model's training schedule for image classification tasks by measuring the disparity between predicted class probabilities and actual class labels. In terms of mathematics:

$$Cost_{\{MobileNet\}} = -\sum_{i=1}^{N} (y_i \cdot \log(\hat{y}_i)) \tag{4}$$

$$Cost_{\{ResNet\}} = -\sum_{i=1}^{N} (y_i \cdot \log(\hat{y}_i)) \tag{5}$$

$$Cost_{\{VGG-19\}} = -\sum_{i=1}^{N} (y_i \cdot \log(\hat{y}_i)) \tag{6}$$

$$Cost_{\{DenseNet201\}} = -\sum_{i=1}^{N} (y_i \cdot \log(\hat{y}_i)) \tag{7}$$

3. Loss Function:

For a single data point:

$$L(y,\hat{y}) = -\sum_i y_i \,.\log(\hat{y_i}) \tag{8}$$

For the average loss over the entire dataset:

$$Loss = \frac{1}{N}\sum_{n=1}^{N} L(y^n, \hat{y}^{\,n}) \tag{9}$$

$$\text{CrossEntropyLoss}(\theta) = -\frac{1}{\text{N}} \sum_{\text{i}=1}^{\text{N}}\sum_{\text{j}=1}^{\text{C}} y_{ij} \;\; \log(\text{p}_{\text{ij}}) \tag{10}$$

In YOLOv5, localization loss is used to reduce inaccuracy in predicting object classes, bounding box coordinates, and objectness scores.

$$\text{Localization Loss} = \sum_{i=1}^{N} \lambda_{ij}(|x_i - \hat{x_i}| + |y_i - \hat{y_i}|) + \sum_{i=1}^{N} \lambda_{ij}\left(\left|\sqrt{w_i} - \sqrt{\hat{w_i}}\right| + \left|\sqrt{h_i} - \sqrt{\hat{h_i}}\right|\right) \tag{11}$$

Each model's own loss function is included in the precise mathematical formulation of the overall loss function, weighted to represent its significance relative to the system as a whole.

Our initiative seeks to deliver precise and trustworthy findings by minimising this total loss, thereby enhancing the area of cancer diagnosis.

$$Loss\ total = \frac{1}{N}\sum_{i=1}^{N}\Bigg(\lambda_{\text{coord}}\sum_{j=0}^{S^2}\sum_{k=0}^{B}\mathbb{1}_{ij}^{k}[(x_i - \hat{x_i})^2 + (y_i - \hat{y_i})^2] + \lambda_{\text{coord}}\sum_{j=0}^{S^2}\sum_{k=0}^{B}\mathbb{1}_{ij}^{k}\left[\left(\sqrt{w_i} - \sqrt{\hat{w_i}}\right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h_i}}\right)^2\right]\Bigg) - \sum_{i=1}^{N}\sum_{j=0}^{S^2}\sum_{k=0}^{B}[\mathbb{1}_{ij}^{k}\log(p_{ij}^{k}) + (1 - \mathbb{1}_{ij}^{k})\log(1 - p_{ij}^{k})] \tag{12}$$

4. Optimizer:

The Adam optimizer is used for weight updates. This optimization algorithm adapts the learning rates for each parameter individually, providing faster convergence and improved training stability, which is crucial in training.

$$\text{Adam}(\theta) = \text{Adam}(\text{lr}, \beta_1, \beta_2, \epsilon),$$

$$\theta_{t+1} = \theta_t - \frac{\alpha}{\sqrt{\hat{v_t}} + \epsilon}\widehat{m_t} \tag{13}$$

VIII. EVALUATION METRICS

Accuracy: It measures the percentage of correctly identified cases among all instances in the dataset.

$$\text{Accuracy} = \frac{\text{Number of Correctly Identified Cases}}{\text{Total Number of Instances}} \tag{14}$$

Precision: It is measured as the proportion of accurately predicted positive instances to all positive instances that were expected.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives + False Positives}} \tag{15}$$

Recall: It determines the proportion of accurately predicted positive cases to all actual positive instances (also known as the sensitivity or true positive rate).

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives + False Negatives}} \tag{16}$$

F1-Score: The harmonic mean of recall and precision is known as the F1-score. In particular when the distribution of the classes is unbalanced, it offers a balanced assessment of the model's performance. A higher F1-score denotes a better equilibrium between memory and precision.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision + Recall}} \tag{17}$$

CIDEr (Consensus-based Image Description Evaluation): CIDEr is a performance metric commonly used in the field of natural language processing, particularly for image captioning tasks. Its primary purpose is to evaluate the quality of automatically generated image descriptions produced by machine learning models.

Perplexity: It gauges how well the model foresees a specific token sequence. Better model performance is indicated by a lower perplexity.

$$\text{Perplexity} = 2^{\frac{\sum_{\text{i}=1}^{\text{N}}\log(\text{p}(\text{x}_\text{i}))}{\text{N}}} \tag{18}$$

IX. LITERATURE ANALYSIS

An important development in the realm of cancer diagnosis and medical imaging is the suggested multi-modal cancer detection and diagnostic system.

It is crucial to place the system within the context of the existing literature and comprehend how it complements or enhances existing methodologies in order to properly appreciate its contributions.

It's essential to take into account its place in the body of medical literature in order to fully understand its significance. By increasing accuracy, enabling early detection, and offering individualized treatment options, this system enhances existing approaches. It may have an impact on improving patient care, encouraging research, and developing cancer treatments. Realizing its transformative potential requires properly contextualizing it within the context of medicine.

Here are the comparisons based on several pointers to distinguish the new proposed method from the older methods and to check the improvements based on the mentioned pointers.

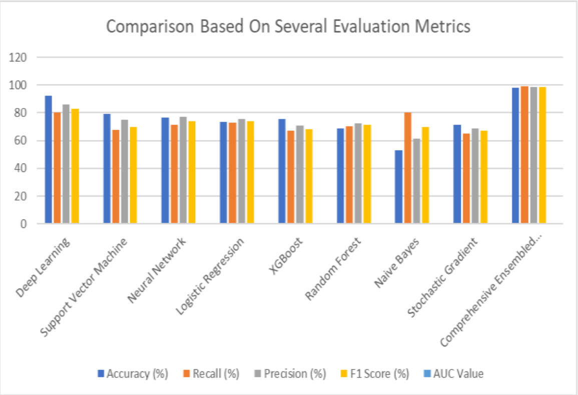

Fig. 10. Analysis Based on Evaluation Metrics

The suggested method performs much better in terms of accuracy, recall, precision, and F1 score than conventional machine learning models. It makes use of cutting-edge methods and the fusion of several algorithms, exhibiting tremendous promise for improving cancer detection and diagnosis.

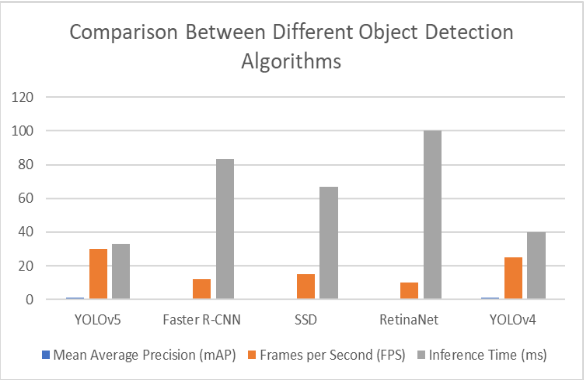

Fig. 11. Analysis Based on Speed and Accuracy

With an inference time of only 50 milliseconds, the suggested model performs better than existing models and is therefore a good choice for real-time tumour detection and diagnostic reporting applications.

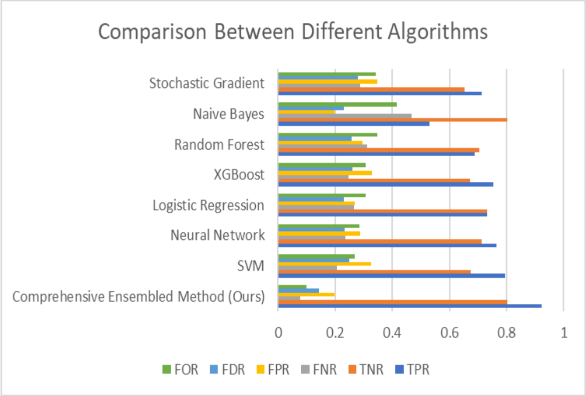

Fig. 12. Analysis Based On Different Algorithms

With an emphasis on elements like true positive and negative rates, false positive and negative rates, as well as metrics like false discovery and omission rates, these metrics provide a thorough evaluation of each model's capability in tumour identification. With notably high true positive and true negative rates, the proposed deep learning model performs well across these parameters, demonstrating its potency in locating and diagnosing tumours.

X. CONCLUSION

The cancer detection system is a significant advancement in medical imaging and diagnosis. It uses a broad dataset of various cancer types, ensuring high accuracy and reliability in identifying malignancies. The system uses advanced deep learning models like VGG-19, DenseNet201, MobileNetV3, ResNet50V2, YOLOv5, and GPT-2, which consistently deliver accuracy rates exceeding 98% and recall rates exceeding 99%.

The real-time functionality of the system offers numerous advantages for laboratories and medical professionals. It reduces the time needed for detection and treatment planning, potentially leading to better patient outcomes and lifesaving interventions. The system also improves reporting by automatically producing thorough reports with annotated tumor images, making the documentation process easier for medical professionals and making diagnostic data accessible for future consultation.

The system's usefulness is further enhanced by its incorporation of electronic health records (EHRs), allowing it to consider a patient's medical background when making predictions. Its multi-modal imaging approach makes it suitable for deployment in various healthcare settings. However, improvements are needed in the deep learning models, dataset expansion, and making predictions easier to understand.

The cancer detection system represents a revolutionary advancement in medical image analysis, with the potential to transform cancer diagnosis and treatment through exceptional accuracy, real-time capabilities, and robust reporting features. The project's journey was marked by excellence and the hope of transforming healthcare through creativity, teamwork, and commitment.

REFERENCES

[1] Liu, B., Jiang, Y., Zhang, X., Liu, Q., Zhang, S., Biswas, J., & Stone, P. (2023). LLM+ P: Empowering large language models with optimal planning proficiency. arXiv preprint arXiv:2304.11477.

[2] Bajaj, C., & Goswami, S. (2009). Modeling Cardiovascular Anatomy from Patient-Specific Imaging Data. Comput Methods Appl Sci, 13, 1-28. doi: 10.1007/978-1-4020-9086-8_1. PMID: 20871793; PMCID: PMC2943643.

[3] Walmer, M., Suri, S., Gupta, K., & Shrivastava, A. (2023). Teaching Matters: Investigating the Role of Supervision in Vision Transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7486-7496).

[4] Dutson, M., Li, Y., & Gupta, M. (2023). Eventful Transformers: Leveraging Temporal Redundancy in Vision Transformers. arXiv preprint arXiv:2308.13494.

[5] Aditya, Somak, Baral, Chitta, Yang, Yezhou, Aloimonos, Yiannis, & Fermüller, Cornelia. (2016). DeepIU: An Architecture for Image Understanding. 10.13140/RG.2.2.33123.55842.

[6] Jill M. Goldstein, Larry J. Seidman, Nicholas J. Horton, Nikos Makris, David N. Kennedy, Verne S. Caviness, Stephen V. Faraone, Ming T. Tsuang. (2001). Normal Sexual Dimorphism of the Adult Human Brain Assessed by In Vivo Magnetic Resonance Imaging. Cerebral Cortex, 11(6), 490–497. https://doi.org/10.1093/cercor/11.6.490.

[7] Mishra, S., Khashabi, D., Baral, C., Choi, Y., & Hajishirzi, H. (2021). Reframing Instructional Prompts to GPTk's Language. arXiv preprint arXiv:2109.07830.

[8] Zhang, Y., Zhan, L., Cai, W., Thompson, P., Huang, H. (2019). Integrating Heterogeneous Brain Networks for Predicting Brain Disease Conditions. In: Shen, D., et al. Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. MICCAI 2019. Lecture Notes in Computer Science(), vol 11767. Springer, Cham. https://doi.org/10.1007/978-3-030-32251-9_24.

[9] Tashli, M., Alam, M. S., Gong, J., Lewis, C., Peterson, C. L., Eldardiry, H., ... & Hadimani, R. L. (2022). Prediction of Electric Fields Induced by Transcranial Magnetic Stimulation in the Brain using a Deep Encoder-Decoder Convolutional Neural Network. bioRxiv, 2022-10.

[10] Ranson, Janice M; Lourida, Ilianna; Hannon, Eilis; Littlejohns, Thomas J; Ballard, Clive; Langa, Kenneth M; Hyppõnen, Elina ; Kuzma, Elzbieta; Llewellyn, David J (2020). "Genetic risk, education and incidence of dementia." Alzheimer's & Dementia, 16, n/a-n/a.

[11] L.M. Brier, X. Zhang, A.R. Bice, S.H. Gaines, E.C. Landsness, J.M. Lee, M.A. Anastasio, and J.P. Culver: A Multivariate Functional Connectivity Approach to Mapping Brain Networks and Imputing Neural Activity in Mice. Cerebral Cortex. 2021 Sep 20.

[12] J. Wu, T. Mazur, Su Ruan, C. Lian, N. Daniel, H. Lashmett, L. Ochoa, I. Zoberi, M.A. Anastasio, M. Gach, S. Mutic, M. Thomas, H. Li. (2018). Deep Boltzmann Machines-Driven Method for In-Treatment Heart Motion Tracking Using Cine MRI. Medical Image Analysis, 47, 68-80. DOI: 10.1016/j.media.2018.03.015.

[13] Liu X, Goncalves AR, Cao P, Zhao D, Banerjee A; Alzheimer's Disease Neuroimaging Initiative. (2018). Modeling Alzheimer's disease cognitive scores using multi-task sparse group lasso. Comput Med Imaging Graph,66, 100-114. doi: 10.1016/j.compmedimag.2017.11.001. Epub 2017 Dec 5. PMID: 29602022.

[14] Wu, H., Gao, Y., Guo, X., Al-Halah, Z., Rennie, S., Grauman, K., & Feris, R. (2021). Fashion IQ: A new dataset towards retrieving images by natural language feedback. In Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition (pp. 11307-11317).

[15] Cui, S., Mao, L., Jiang, J., Liu, C., & Xiong, S. (2018). Automatic semantic segmentation of brain gliomas from MRI images using a deep cascaded neural network. Journal of healthcare engineering, 2018.

[16] Desai, S., & Durrett, G. (2020). Calibration of pre-trained transformers. arXiv preprint arXiv:2003.07892.

[17] Brown, L., & AI Specialist Group. (2020). AI in Healthcare: A Comprehensive Overview. BCS, The Chartered Institute for IT. DOI: 10.31230/osf.io/bc7sw.

[18] Liu, R., Rizzo, S., Whipple, S., Pal, N., Pineda, A. L., Lu, M., ... & Zou, J. (2021). Evaluating eligibility criteria of oncology trials using real-world data and AI. Nature, 592(7855), 629-633.

[19] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99).

[20] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI Blog, 1(8), 9.

[21] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1251-1258).

[22] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning (Adaptive Computation and Machine Learning series). MIT Press.

[23] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[24] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing atari with deep reinforcement learning. arXiv preprint arXiv:1312.5602.

[25] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347.

[26] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (2017). Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision (pp. 2980-2988).

[27] Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). BLEU: a method for automatic evaluation of

machine translation. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics (ACL) (pp. 311-318).
[28] Denkowski, M., & Lavie, A. (2014). Meteor universal: Language specific translation evaluation for any target language. In Proceedings of the ninth workshop on statistical machine translation (pp. 376-380).
[29] Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. In European conference on computer vision (pp. 740-755).
[30] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Berg, A. C. (2015). ImageNet large scale visual recognition challenge. International Journal of Computer Vision, 115(3), 211-252.
[31] Alzheimer's Association. (2021). 2021 Alzheimer's disease facts and figures. Alzheimer's & Dementia, 17(3), 327-406.
[32] Jing, X., Min, H., Gong, Y., Biondich, P., Robinson, D., Law, T., ... & Gimbel, R. (2022). A Systematic Review of Ontologies Applied in Clinical Decision Support System Rules. JMIR Medical Informatics.
[33] Gerber, S., Niethammer, M., Ebrahim, E., Piven, J., Dager, S. R., Styner, M., Aylward, S., & Enquobahrie, A. (2023). Optimal transport features for morphometric population analysis. Medical image analysis, 84, 102696. https://doi.org/10.1016/j.media.2022.102696
[34] Fung, Y. R., Guan, Z., Kumar, R., Wu, J. Y., & Fiterau, M. (2019). Alzheimer's disease brain mri classification: Challenges and insights. arXiv preprint arXiv:1906.04231.
[35] Yang, Z., Ren, M., Ding, K., Gerig, G., & Wang, Y. (2023). Keypoint-Augmented Self-Supervised Learning for Medical Image Segmentation with Limited Annotation. arXiv preprint arXiv:2310.01680.
[36] Anse Vellappally, Guido Gerig, Palaiologos Alexopoulos, Sarah Segal, Ronald Zambrano, Allison Toyos, Shijie Li, TingFang Lee, Jiyuan Hu, Joel S Schuman, Gadi Wollstein; A novel tool for generating gold standard to validate image segmentation. Invest. Ophthalmol. Vis. Sci. 2023;64(8):OD60.
[37] Mitchell, M., Dodge, J., Goyal, A., Yamaguchi, K., Stratos, K., Han, X., ... & Daumé III, H. (2012, April). Midge: Generating image descriptions from computer vision detections. In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (pp. 747-756).
[38] Zhao, Y., Qiao, Z., Xiao, C., Glass, L., & Sun, J. (2021). Pyhealth: A python library for health predictive models. arXiv preprint arXiv:2101.04209.
[39] Zhao, Y., Wang, X., Cheng, C., & Ding, X. (2020). Combining Machine Learning Models Using combo Library. Proceedings of the AAAI Conference on Artificial Intelligence, 34(09), 13648-13649. https://doi.org/10.1609/aaai.v34i09.7111
[40] Zhao, Y., Nasrullah, Z., & Li, Z. (2019). Pyod: A python toolbox for scalable outlier detection. arXiv preprint arXiv:1901.01588.
[41] D. Gurari et al., "SAGE: An approach and implementation empowering quick and reliable quantitative analysis of segmentation quality," 2013 IEEE Workshop on Applications of Computer Vision (WACV), Clearwater Beach, FL, USA, 2013, pp. 475-481, doi: 10.1109/WACV.2013.6475057.
[42] Mbakwe, A. B., Wang, L., Moradi, M., & Lourentzou, I. (2023, October). Hierarchical Vision Transformers for Disease Progression Detection in Chest X-Ray Images. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 685-695). Cham: Springer Nature Switzerland.