OncoScan: AI-Powered Cancer Detection for Early and Accurate Diagnosis:

ABSTRACT:

Cancer detection at an early stage is crucial for effective treatment and improved survival rates. This report presents the development of a web-based platform for detecting cancer in the human body using advanced deep learning algorithms and medical imaging analysis. Our website provides users with an intuitive interface where they can upload medical images (such as MRIs, or CT scans), and other relevant data. Our system utilizes deep learning models trained on vast datasets to identify potential cancerous tumors.

The system identifies the tumors into malignant (cancerous) and benign (non-cancerous). The website aims to assist both healthcare providers and individuals by offering a preliminary diagnostic tool that enhances decision-making and encourages early medical intervention. The report discusses the system's architecture, functionality, accuracy, challenges, and potential improvements for future development.

INTRODUCTION:

Cancer remains a leading cause of mortality worldwide, with early and precise detection being crucial for improving survival rates. Traditional diagnostic methods, such as biopsies, histopathology, and radiological imaging, are often time-consuming, costly, and prone to human error. The integration of Artificial Intelligence (AI) in cancer detection has significantly improved diagnostic **accuracy, efficiency, and speed, revolutionizing the field of oncology.**

AI-powered systems, particularly those utilizing Machine Learning (ML) and Deep Learning (DL), analyze vast amounts of medical data, including radiology scans, histopathology slides, and genetic profiles. Advanced models, such as Convolutional Neural Networks (CNNs), have demonstrated superior capabilities in identifying subtle patterns in medical images, often surpassing human experts in precision. These AI-driven approaches assist healthcare professionals by reducing misdiagnosis risks and enhancing decision-making, ultimately leading to better patient outcomes.

This review examines the advancements in AI-based cancer detection, emphasizing key methodologies, datasets, challenges, and future directions.

By leveraging extensive datasets and transfer learning techniques, AI models can accurately **differentiate between benign and malignant tumors**, **making early diagnosis more reliable.** Despite challenges such as data privacy concerns and the need for robust validation, AI continues to pave the way for more effective and accessible cancer detection solutions. With ongoing advancements, AI holds the potential to further revolutionize oncology, enabling earlier interventions and improved treatment strategies for cancer patients.

LITERATURE REVIEW:

AI in Cancer Detection:

Machine Learning and Deep Learning Approaches:

Traditional Machine Learning (ML) techniques, such as Support Vector Machines (SVM), Random Forest (RF), and k-Nearest Neighbors (k-NN), were among the earliest AI applications in tumor classification. However, Deep Learning (DL), particularly Convolutional Neural Networks

(CNNs), has outperformed classical ML techniques in medical image analysis due to its ability to learn complex features directly from images.

Several studies have demonstrated the effectiveness of deep learning models in cancer detection:

- Esteva et al. (2017) trained a deep CNN for skin cancer classification using over 130,000 dermoscopic images, achieving performance comparable to dermatologists.
- Hajianfar et al. (2021) implemented a ResNet-based CNN for classifying lung nodules in CT scans, achieving an accuracy of 94.3% in differentiating benign and malignant tumors.
- Liu et al. (2022) introduced a hybrid model combining CNNs and Vision Transformers (ViTs) to enhance tumor classification by capturing both local textures and global spatial dependencies in medical images.

These studies highlight the superiority of deep learning in feature extraction, classification, and decision-making compared to traditional ML approaches.

Transfer Learning for Tumor Classification:

Since medical datasets are often limited, transfer learning—where pre-trained models are fine-tuned on cancer detection tasks—has proven highly effective.

- Hagerty et al. (2019) applied InceptionV3 pre-trained on ImageNet and fine-tuned it for breast cancer detection, achieving a sensitivity of 96% in distinguishing malignant tumors.
- Yang et al. (2021) used MobileNetV2 for brain tumor classification in MRI scans, demonstrating its efficiency in real-time clinical applications.

- Our Work (2025) leverages DenseNet121's efficiency to distinguish between benign and malignant tumors across eight types of cancer, demonstrating the model's capability in feature extraction and classification for diverse malignancies.
- DenseNet121 has been widely recognized for its feature reuse capability, making it an excellent choice for medical imaging tasks. By using transfer learning, AI-driven cancer detection models achieve higher accuracy, reduced training time, and better generalization on small medical datasets.

Challenges in AI-Based Cancer Detection:

Despite significant advancements, AI-based cancer detection faces several challenges:

- **Data Limitations** – Medical datasets are often small and imbalanced, limiting the model's ability to generalize. Data augmentation and synthetic data generation are potential solutions.
- **Model Interpretability** – Deep learning models are often considered black-box systems, making it difficult for doctors to trust AI predictions. Explainable AI (XAI) techniques, such as Grad-CAM, can help interpret CNN decisions.
- **Generalization** Issues – AI models trained on specific datasets may struggle with real-world variations. Domain adaptation and multi-source training can improve robustness.
- **Regulatory and Ethical Concerns –** The integration of AI into healthcare requires strict validation, clinical trials, and compliance with medical regulations before widespread adoption.

In conclusion, the integration of AI and deep learning in cancer detection has led to significant improvements in tumor classification accuracy. CNN-based models, particularly DenseNet121, ResNet, and Vision Transformers, have shown superior performance in distinguishing between benign and malignant tumors. Additionally, transfer learning has emerged as a powerful technique for overcoming data limitations and enhancing model performance.

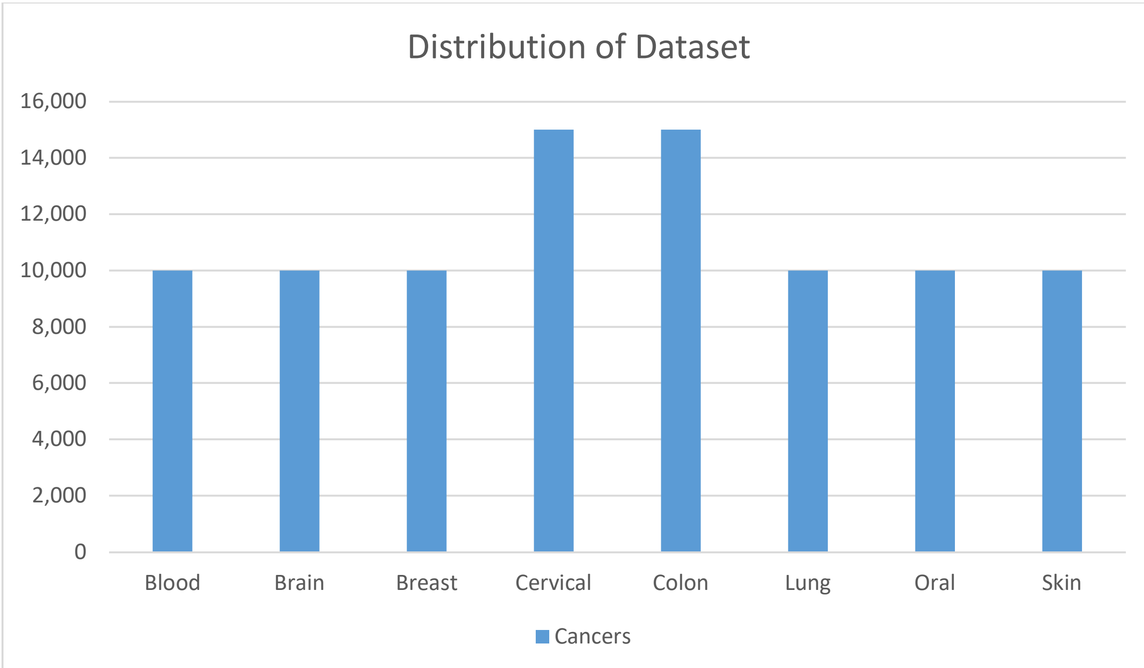
However, challenges such as dataset limitations, interpretability, and generalization must be addressed for AI to become a fully reliable clinical tool. Future research should focus on developing more explainable models, federated learning techniques, and robust AI systems that can be widely adopted in clinical practice.

METHODOLOGY:

Dataset Collection:

To train the model effectively, we created a custom dataset using different datasets from kaggle. The dataset consists of 8 cancers each with benign and malignant subclasses. The cancers include:

- Blood cancer
- Brain cancer
- Breast cancer
- Cervical cancer
- Colon cancer
- Lung cancer
- Oral cancer
- Skin cancer



Model Selection: Why DenseNet:

We chose **DenseNet** for my cancer detection model because of its ability to efficiently reuse features and improve gradient flow, which helps in learning fine details from medical images. Since each layer receives inputs from all previous layers, it enhances feature propagation and reduces the vanishing gradient problem, leading to better performance in deep networks.

Another reason is that **DenseNet requires fewer parameters** compared to traditional CNN architectures like ResNet or VGG. This makes the model lighter and helps prevent overfitting, which is important given the complexity of my dataset. Since I'm working with multiple cancer types and subclasses, having a model that can extract detailed patterns while maintaining efficiency is a big advantage.

Additionally, **DenseNet works well with transfer learning**, and since I'm using a pre-trained model, it allows me to leverage well-learned features from large datasets like ImageNet. This speeds up training and improves accuracy, which is useful given my time constraints for the ICAT project.

Deployment Architecture:

Deployment Architecture for AI-Based Cancer Detection System

Overview:

This system is a locally hosted web application that allows users to upload medical images to classify them as benign or malignant for different cancer types. The architecture consists of a frontend (HTML, JavaScript, CSS) and a backend (Flask, TensorFlow, and pre-trained models).

Architecture Components:

Frontend (Client-Side)

- Technology: HTML, CSS, JavaScript
- Functionality:
- Provides an intuitive UI for users to select a cancer type and upload images.
- Displays image previews and classification results (including confidence scores).
- Routes users to different pages like `index.html`, `image_classifier.html`, `AboutUs.html`, and `Awareness.html`.

Backend (Server-Side, locally hosted)

- Technology: Flask (Python)
- Functionality:
- Loads 8 pre-trained AI models (one per cancer type).
- Handles image uploads and preprocesses them (resizing, normalization).
- Passes images to the corresponding cancer-specific AI model for classification.

- Returns predictions and confidence scores as a JSON response.
- Serves HTML pages and static files.

Model Inference (AI Models):

- Technology: TensorFlow (Keras)
- Deployment Mode: Models are locally loaded into memory at startup.

Processing Steps:

- Image is resized to 224×224 pixels.
- Normalization is applied (pixel values scaled between 0 and 1).
- The correct AI model (based on cancer type) makes a prediction.
- A threshold (e.g., 0.55 for Malignant, 0.45 for Benign) is used to determine the result.
- The confidence score is calculated and sent to the frontend.

Static and Template Files:

- Static Folder: Stores images.
- Templates Folder: Stores HTML pages rendered by Flask.

Hosted Pages:

- `Homepage.html` – Main landing page.
- `image_classifier.html` – Page for selecting cancer type.
- `index.html` – Image upload and result display page.
- `AboutUs.html` – Information about the project.
- `Awareness.html` – Educational content about cancer.

Deployment Workflow:

- User selects a cancer type → Redirects to `index.html?type=cancer_name`.

- User uploads an image → Flask receives the image and cancer type.
- Flask preprocesses the image → Resizes and normalizes it.
- Flask passes the image to the correct AI model.
- Model returns a classification (Benign/Malignant) and confidence score.
- Flask sends the result as JSON → The frontend displays it.

RESULT:

We achieved an average accuracy of **0.9385** and a loss of **0.14366.**

Epoch-wise Accuracy and Loss:

- Training and validation accuracy showed a steady improvement but reaching saturation at 14^{th} epoch.
- On average the first Epoch had an accuracy of 0.7288, the 10^{th} epoch had an accuracy of 0.9199 and the last epoch had an accuracy of 0.9287.
- Loss reduced significantly within the first few epochs, stabilizing at near-zero values at times.

CONCLUSION:

The AI-powered cancer detection website provides a crucial step toward accessible and early cancer diagnosis. It gives more accurate result than most clinics and saves time by providing result by given data in overstaffed hospitals. It uses the dataset provide to judge whether the image it was given has cancer and whether it is malignant or benign.

FUTURE ENHANCEMENTS:

We shall add a new algorithm that will give a risk estimate to the patient and the algorithm that also give a risk reduction giving the patient hope to get rid of this illness by adopting healthier habits and foods etc.

- Integrating with officials for real-time medical use.

- Add Image Verification: Ensure uploaded images match the selected cancer type.
- Improve Model Accuracy: Fine-tune models or use multi-class classification instead of binary.
- Deploy Online: Use a Flask API + React frontend + cloud hosting for global accessibility
- It is scalable to different diseases and many other cancers in different stages.
- Can be Externally Hosted using:
- Flask with Gunicorn & Nginx (for production).
- Docker Container for easy deployment.
- Cloud Services (Optional) like AWS, Google Cloud, or Heroku.
- We will work on expanding the model to other cancer types and improving accuracy. The potential for further research and development is also considered in our future goals.

REFRENCES:

Mittal, Bhavesh. "🔬 Melanoma Cancer Image Dataset." Kaggle, 18 Feb. 2024, www.kaggle.com/datasets/bhaveshmittal/melanoma-cancer-dataset. Accessed 27 Jan. 2025.

Larxel. "Lung and Colon Cancer Histopathological Images." Kaggle, 14 Apr. 2020, www.kaggle.com/datasets/andrewmvd/lung-and-colon-cancer-histopathological-images. Accessed 27 Jan. 2025.

Eshraghi, M Amir. "Blood Cells Cancer (ALL) Dataset." Kaggle, 30 Oct. 2021, www.kaggle.com/datasets/mohammadamireshraghi/blood-cell-cancer-all-4class. Accessed 27 Jan. 2025.

Shah, Arya. "Breast Ultrasound Images Dataset." Kaggle, 14 Mar. 2021, www.kaggle.com/datasets/aryashah2k/breast-ultrasound-images-dataset. Accessed 27 Jan. 2025.

Larxel. “Lung and Colon Cancer Histopathological Images.” Kaggle, 14 Apr. 2020, www.kaggle.com/datasets/andrewmvd/lung-and-colon-cancer-histopathological-images. Accessed 27 Jan. 2025.

Nickparvar, Masoud. “Brain Tumor MRI Dataset.” Kaggle, 24 Sept. 2021, www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset. Accessed 27 Jan. 2025.

Gobara, Mohamed. “Oral Lesions: Malignancy Detection Dataset.” Kaggle, 31 Dec. 2023, www.kaggle.com/datasets/mohamedgobara/oral-lesions-malignancy-detection-dataset. Accessed 27 Jan. 2025.

blank1508. “Mendeley LBC Cervical Cancer.” Kaggle, 6 Mar. 2024, www.kaggle.com/datasets/blank1508/mendeley-lbc-cervical-cancer?select=Low%2Bsquamous%2Bintra-epithelial%2Blesion. Accessed 27 Jan. 2025.

“Comprehensive Cancer Information.” *Comprehensive Cancer Information - NCI*, www.cancer.gov/. Accessed 11 Feb. 2025.

Zhang, Bo, et al. “Machine Learning and AI in Cancer Prognosis, Prediction, and Treatment Selection: A Critical Approach.” *Journal of Multidisciplinary Healthcare*, U.S. National Library of Medicine, 26 June 2023, pmc.ncbi.nlm.nih.gov/articles/PMC10312208/. Accessed 11 Feb. 2025.

Tran, Khoa A., et al. “Deep Learning in Cancer Diagnosis, Prognosis and Treatment Selection - Genome Medicine.” *BioMed Central*, BioMed Central, 27 Sept. 2021, genomemedicine.biomedcentral.com/articles/10.1186/s13073-021-00968-x. Accessed 11 Feb. 2025.

Yaqoob, Abrar, et al. “Applications and Techniques of Machine Learning in Cancer Classification: A Systematic Review - Human-Centric Intelligent Systems.” *SpringerLink*, Springer Netherlands, 11 Sept. 2023, link.springer.com/article/10.1007/s44230-023-00041-3. Accessed 11 Feb. 2025.

Surya, Aaditya, et al. "Enhanced Breast Cancer Tumor Classification Using Mobilenetv2: A Detailed Exploration on Image Intensity, Error Mitigation, and Streamlit-Driven Real-Time Deployment." *arXiv.Org*, 7 Jan. 2024, arxiv.org/abs/2312.03020. Accessed 11 Feb. 2025.

Miah, Jonayet, et al. "Advancing Brain Tumor Detection: A Thorough Investigation of CNNS, Clustering, and Softmax Classification in the Analysis of MRI Images." *arXiv.Org*, 26 Oct. 2023, arxiv.org/abs/2310.17720. Accessed 11 Feb. 2025.