A NOVEL LUNG CANCER DETECTION SYSTEM USING DEEP LEARNING

V.MANISHA[1],M. HARIKA[1],P. BHUVANESVAR[1],B. SIVA SANKAR[1],S. SRIKANTH REDDY[1]

J. SIVA PARVATHI[2]

B-Tech Student, Dept.of CSE,UNIVESAL COLLEGE OF ENGINEERING AND TECHNOLOGY,Andhra pradhesh, India [1]

Assist Professor, Dept.of CSE,UNIVERSAL COLLEGE OF ENGINEERING AND TECHNOLOGY,Andhra pradhesh, India[2]

manishavalluri7@gmail.com , shivaparvathi@gmail.com

ABSTRACT

Lung cancer remains one of the most prevalent and deadly diseases globally, with early diagnosis being pivotal for improving patient outcomes. Detecting lung cancer at an early stage can significantly enhance survival rates, but this often requires specialized diagnostic tools and expertise. To address this challenge, this project develops a Django-based web application that integrates a deep learning model for lung cancer detection. The system utilizes Convolutional Neural Networks (CNNs), a powerful type of deep learning model known for their exceptional ability to process and classify visual data, particularly medical imaging such as X-rays and CT scans. The deep learning model, developed using TensorFlow/Keras, is trained to recognize patterns in medical images and classify them into two categories: images with lung cancer and images without lung cancer. This model is continuously trained on a curated dataset of annotated medical images, and it adapts to new data to improve prediction accuracy over time. The web application provides a comprehensive platform where medical professionals and researchers can upload new medical images, enabling real-time prediction of lung cancer presence.The user interface is designed to be intuitive and accessible, allowing users to interact with the system without requiring advanced technical expertise. Key features include image upload functionality, model training capabilities, and a real- time prediction feature. Users can monitor the model's training progress through visualizations of training accuracy and loss graphs, which help track the model's performance over epochs.By integrating CNNs with a Django web application, this project creates a robust tool that aids in the automated detection of lung cancer, improving diagnostic workflows, enhancing early detection, and providing a resource for ongoing research and development in medical image processing. The application is designed with scalability in mind, allowing for continuous improvement as more medical data becomes available.

Keywords: Lung cancer, Deep learning, CNN, Medical imaging, Django, TensorFlow, Keras, Early detection, Image classification, Real-time prediction, Web application.

1. INTRODUCTION

Lung cancer is one of the leading causes of cancer-related deaths worldwide, accounting for millions of fatalities annually. The disease is often diagnosed at advanced stages, where treatment options are limited, and survival rates are low. Early detection of lung cancer, however, significantly improves the chances of successful treatment and patient survival. Despite advancements in diagnostic techniques, the process remains complex, requiring specialized expertise and often leading to delayed diagnoses.Medical imaging, particularly chest X-rays and CT scans, plays a crucial role in the early detection of lung cancer. However, manual interpretation of these images can be time-consuming, error-prone, and highly dependent on the skill of the radiologist. With the growing volume of medical data, there is an increasing need for automated systems that can assist healthcare professionals in making accurate and timely diagnoses.This project proposes a solution by integrating deep learning-based image classification into a web application built with Django. The core of the system is a Convolutional Neural Network (CNN), a class of deep learning algorithms that have proven highly effective in analyzing medical images for the detection of various conditions, including lung cancer. By training a CNN on large datasets of annotated X-ray and CT scan images, the model learns to recognize patterns indicative of lung cancer, offering a reliable, fast, and scalable method for detecting the disease.The Django-based web application serves as the interface for interacting with the deep learning model, allowing users to upload medical images, train the model, and obtain real- time predictions. Additionally, the application provides essential visualizations, including graphs that track training accuracy and loss, helping users monitor the model's performance and make informed decisions about further training.The main objective of this project is to bridge the gap between cutting-edge deep learning techniques and real-world medical applications. By automating the image analysis process, the system aims to reduce the burden on healthcare professionals, increase diagnostic efficiency, and ultimately improve patient outcomes by enabling earlier detection of lung cancer.

This paper outlines the development of this system, starting from the pre-processing of medical image data to the training and deployment of the deep learning model.

2. LITERATURE SURVEY

Lung cancer is one of the leading causes of cancer-related deaths worldwide. Early and accurate diagnosis is critical to improving patient survival rates. Traditional diagnostic methods such as biopsy and radiological imaging are time-consuming and sometimes yield ambiguous results. With advancements in artificial intelligence, especially deep learning, automated and accurate lung cancer detection has become more viable and reliable. This literature survey reviews key studies and systems that have contributed to the development of deep learning-based lung cancer detection techniques.

2.1. Medical Imaging and Deep Learning

Medical imaging, especially **Computed Tomography (CT)** and **X-rays**, plays a pivotal role in lung cancer detection. A study by **Setio et al. (2016)** introduced a multi-view CNN approach to detect lung nodules in CT scans. This work emphasized the importance of spatial and contextual information in nodule classification, which is incorporated in many current deep learning models.

2.2. Convolutional Neural Networks (CNNs)

CNNs are widely used in image classification tasks. Research by **Shen et al. (2017)** applied deep CNNs to classify lung nodules as benign or malignant using 3D CT data. The model demonstrated promising results and confirmed that deeper networks can capture high-level features essential for accurate cancer detection.

2.3. Transfer Learning

Transfer learning has proven effective in improving performance, especially when dealing with limited medical data. **Rajpurkar et al. (2017)** used a pre-trained DenseNet model on chest X-rays and achieved near radiologist-level accuracy in detecting pneumonia, a method also adapted in recent lung cancer detection systems to enhance performance with smaller datasets.

2.4. LIDC-IDRI Dataset

The **Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI)** is a commonly used benchmark dataset in lung cancer research. It provides labeled CT scans with annotations from multiple radiologists. Many deep learning models, including those proposed by **Armato et al. (2015)**, have been trained and tested on this dataset to validate their accuracy and robustness.

2.5. Segmentation and ROI Extraction

Segmentation of Regions of Interest (ROIs) such as lung nodules is a critical preprocessing step. **UNet** and its variations have been widely used for medical image segmentation. **Ronneberger et al. (2015)** developed the original UNet architecture, which has since been widely adopted in lung cancer imaging studies for precise nodule localization.

2.6. Performance Metrics

Common evaluation metrics include **accuracy, precision, recall, F1-score, and AUC (Area Under Curve)**. These metrics provide a comprehensive understanding of a model's performance. Studies by **Kermany et al. (2018)** and others highlight the need for high sensitivity in cancer detection to minimize false negatives.

2.7. Challenges and Future Scope

Despite the progress, challenges such as data imbalance, variations in tumor size and shape, and false positives remain. Recent approaches suggest combining **CNNs with Recurrent Neural Networks (RNNs)** or **attention**

doi: 10.48047/ijiee.2025.15.4.8

mechanisms to improve temporal and spatial understanding in scan sequences. Additionally, **explainable AI (XAI)** techniques are being developed to enhance clinical trust in AI-based diagnosis systems.

3. EXISTING SYSTEM:

The detection and diagnosis of lung cancer have traditionally relied on various imaging techniques, with chest X-rays and computed tomography (CT) scans being the most commonly used. While these methods have proven effective in identifying abnormalities, they come with their own set of challenges, such as a reliance on the expertise of medical professionals and the potential for human error.

3.1 Chest X-rays:

Chest X-rays are among the most common diagnostic tools used to detect lung cancer. They are widely available, cost-effective, and can quickly identify abnormal masses or signs of lung cancer. However, the effectiveness of chest X- rays is limited by their resolution and inability to detect small tumors, particularly in the early stages of the disease. Moreover, interpreting X-rays requires significant expertise, and there is always a risk of misdiagnosis due to overlapping conditions, such as pneumonia or tuberculosis, which can present similar symptoms on an X-ray.

3.2 Computed Tomography (CT) Scans:

CT scans provide a more detailed view of the lungs compared to X-rays, making them a valuable tool for detecting and staging lung cancer. CT scans can detect smaller tumors and are particularly useful for evaluating the size, location, and spread of cancer. However, CT scans also come with some disadvantages. They involve exposure to higher levels of radiation, which can pose risks if patients undergo multiple scans over time. Furthermore, interpreting CT images require specialized radiologists, and even they may miss subtle signs of early- stage cancer.

3.3 Positron Emission Tomography (PET) Scans:

PET scans are used to assess the metabolic activity of cells, including cancerous cells, in the body. They are particularly helpful in determining whether lung cancer has spread to other parts of the body (metastasis). However, PET scans are expensive, not as widely available, and require additional equipment and expertise. Moreover, while PET scans are useful in staging cancer, they are not typically used for the initial detection of lung cancer.

3.4 Biopsy and Histopathological Examination:

While imaging techniques are essential for detecting lung cancer, a biopsy— removing a tissue sample from the lung—is the most definitive way to confirm the diagnosis. Biopsy samples are examined under a microscope to identify cancer cells. However, biopsy procedures are invasive and carry risks such as bleeding, infection, and puncturing of the lung. Additionally, biopsies are not always feasible if the tumor is located in hard-to-reach areas.

3.5 Manual Image Interpretation:

One of the biggest challenges in lung cancer detection is the manual interpretation of medical images. Radiologists review the images and search for abnormal patterns or masses indicative of cancer. Despite their expertise, manual interpretation can be time-consuming, subject to human error, and influenced by fatigue or workload. Studies have shown that even experienced radiologists may miss up to 30% of small lung tumors, which underscores the need for automated systems to assist in image analysis.

Limitations of Existing Methods:

While the current methods for lung cancer detection have been instrumental in early diagnosis, they are far from perfect. Key limitations include:

- **Human Error:** Interpretation of X-rays, CT scans, and PET scans is prone to human error, even among trained specialists.
- **Time-Consuming:** The process of manual interpretation is time- consuming and may delay diagnosis, especially when large volumes of scans need to be reviewed.
- **Cost and Availability:** Some imaging techniques, such as PET scans and advanced CT scans, are expensive and may not be readily available in all healthcare facilities, particularly in low-resource settings.
- **Radiation Exposure:** CT scans and PET scans expose patients to higher doses of radiation, which may

limit their use, especially in screening high- risk individuals.

- **Invasive Procedures:** Biopsies are often required for definitive diagnosis but involve risks and are not always suitable for every patient.

4. PROPOSED SYSTEM:

The proposed solution integrates deep learning with a web-based platform for lung cancer detection. This approach combines the power of Convolutional Neural Networks (CNNs) with the accessibility and scalability of a Django web application. The aim is to automate and enhance the lung cancer detection process, providing more accurate, efficient, and accessible diagnostics compared to existing methods.

4.1 Convolutional Neural Network (CNN) for Lung Cancer Detection:

At the heart of the proposed system is a CNN model, a type of deep learning architecture that excels in image classification tasks. CNNs automatically learn and extract relevant features from images, allowing them to identify patterns that are indicative of lung cancer. The CNN model will be trained using a large dataset of labelled medical images, such as X- rays and CT scans, where each image is classified as either showing signs of lung cancer or being normal. The model will be trained using TensorFlow/Keras, popular deep learning frameworks that provide a high-level interface for building and training neural networks. The CNN will be designed with multiple convolutional layers that learn hierarchical features (such as edges, textures, and shapes) in the images. Pooling layers will help in reducing the dimensionality of the feature maps, improving computational efficiency, while fully connected layers at the end of the network will make the final prediction of whether the image contains lung cancer or not.

4.2 Web Application (Django-based Interface):

The CNN model will be integrated into a web application built using Django, a popular Python web framework. This platform will allow users to upload medical images (X-rays or CT scans), which will be processed and analyzed by the trained CNN model. The key features of the web application include:

- Image Upload and Prediction: Users can upload medical images directly through the application interface. The images will be pre-processed (e.g., resized and normalized) before being fed into the CNN model for prediction.

- Model Training Interface: The application will also allow users to train the model using new datasets. This feature will be particularly useful for medical institutions or research groups that wish to update or fine-tune the model with their own data. The training process will be visualized with real-time graphs showing the accuracy and loss of the model over epochs.

- Visualization of Metrics: The web application will provide a dashboard that displays training progress, accuracy, loss curves, and other relevant metrics, helping users monitor the model's performance during training and validation. This feature enables transparency in the model's learning process and offers insights into its reliability.

- Real-time Predictions: Once the model is trained, it will be able to classify newly uploaded images in real-time. The results will be presented in a user- friendly format, with additional information such as the probability score of the prediction and confidence intervals.

4.3 Data Preprocessing:

For the CNN model to perform effectively, preprocessing of the medical image data is crucial. Images will be standardized in size (e.g., 224x224 pixels) and normalized to a consistent scale. Image augmentation techniques such as rotation, flipping, and scaling will be applied to increase the diversity of the training data and prevent overfitting. This preprocessing pipeline will ensure that the model can generalize well to unseen images.

4.4 Model Evaluation and Performance:

The proposed method will emphasize a rigorous evaluation of the trained CNN model using performance metrics such as accuracy, precision, recall, F1-score, and the Area Under the Curve (AUC) for the Receiver Operating Characteristic (ROC) curve. These metrics will be crucial for assessing the model's effectiveness in detecting lung cancer and reducing false positives and false negatives, which can be life-threatening in medical applications.

doi: 10.48047/ijiee.2025.15.4.8

The model will also undergo cross-validation to ensure that it performs well across different subsets of the dataset, preventing issues such as overfitting to a particular set of images. Additionally, transfer learning techniques may be employed to leverage pre-trained models (e.g., VGG16 or ResNet) and fine-tune them on the lung cancer detection dataset. This approach will help speed up training and improve the model's performance, especially when dealing with limited labeled data.

4.5 Integration of Machine Learning with Django:

Integrating the deep learning model with the Django web application will be achieved through an API layer, where the model can be deployed as a RESTful service. Users will interact with the frontend via the web interface, while the backend (Django) will handle tasks such as image processing, model inference, and providing real-time results.

The application will be hosted on a scalable cloud platform (e.g., AWS, Heroku, or Google Cloud), ensuring that the system is accessible globally and can handle large numbers of concurrent users without compromising on performance.

4.6 Benefits of the Proposed Method:

- Automation: The deep learning model automates the process of analyzing medical images, reducing the reliance on human expertise and minimizing errors associated with manual interpretation.

- Scalability and Accessibility: The web application is accessible through any device with an internet connection, enabling healthcare providers worldwide to access the tool, regardless of location or resources.

- Real-Time Predictions: The system offers quick and accurate predictions, allowing for faster diagnosis and decision-making.

- Improved Diagnostic Efficiency: By reducing the time spent on manual analysis, healthcare professionals can focus on other critical tasks, improving overall workflow efficiency.

- User-Friendly Interface: The Django-based web application provides an intuitive interface, making it accessible to medical professionals without requiring deep technical knowledge.

4.7 Future Improvements:

- Expansion to Other Diseases: The method can be extended to detect other types of cancer or medical conditions by training the CNN model on relevant datasets.

- Model Fine-Tuning: Continuous improvements in the model can be achieved by integrating more diverse datasets, allowing for better generalization and accuracy across different populations.

- Integration with Electronic Health Records (EHR): In the future, the system can be integrated with hospital databases and EHR systems for seamless diagnosis and patient tracking.

Workflow of the Proposed System:

1 Input: Users provide crime stories either by typing them into the system or using speech recognition to provide verbal descriptions.

2 Processing: The system uses NLP techniques (such as Named Entity Recognition, dependency parsing, and word embeddings) to extract key elements (suspects, motives, and opportunities) and preprocess the narrative for machine learning models.

3 Prediction: The crime type is predicted using machine learning models trained on historical crime data.

4 Analysis: The system generates insights, such as profiling suspects, identifying motives, and understanding crime opportunities. This information is presented through a dashboard, including visual representations of the key elements.

5 Output: The system communicates the results using text-to-speech functionality and displays detailed crime analysis on the user interface.

System Architecture:

doi: 10.48047/ijiee.2025.15.4.8

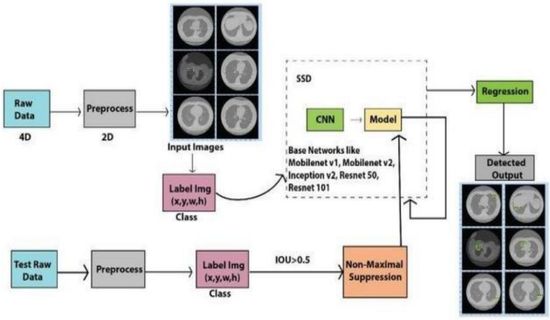

Fig 1.1. system architecture

The system processes raw 4D medical scan data by converting it into 2D images through preprocessing. These images are labeled with object locations and class information. The labeled images are then passed into an SSD (Single Shot Detector) model, which uses base networks like MobileNet or ResNet to detect objects. After detection, overlapping predictions are filtered using Non-Maximal Suppression to keep only the best matches. Finally, the results are refined using regression and shown as detected regions on the scan images.

5 RELATED WORK

5.1. Basic CNN Architecture

A standard CNN architecture consists of multiple convolutional layers followed by pooling layers, activation functions (like ReLU), and fully connected layers. In the context of lung cancer detection, these architectures are trained to classify CT or X-ray images into healthy or cancerous categories. This simple CNN model often serves as a baseline in many studies.

5.2. AlexNet

AlexNet, developed by Krizhevsky et al., was one of the earliest deep CNN architectures to demonstrate excellent image classification performance. It has been adapted for medical imaging by fine-tuning on lung cancer datasets. The use of ReLU activation, dropout layers, and data augmentation improved generalization and reduced overfitting in medical applications.

5.3. VGGNet

VGG16 and VGG19, known for their deep architecture with small (3x3) filters, have been utilized for lung nodule classification. Their uniform architecture enables easy adaptation and transfer learning, making them popular in detecting and classifying lung nodules with limited data availability**.**

5.4. ResNet (Residual Networks)

ResNet, especially ResNet50 and ResNet101, introduced skip connections to address the vanishing gradient problem in deep networks. This architecture has been used in lung cancer detection to achieve high accuracy while maintaining deeper networks. The residual blocks help in better learning of fine-grained features from CT or X-ray scans.

5.5. DenseNet

Dense Convolutional Networks (DenseNet) connect each layer to every other layer in a feed-forward fashion. In lung cancer applications, this dense connectivity helps in better feature propagation and reuse. DenseNet has been shown to outperform traditional CNNs on small and imbalanced datasets common in medical imaging.

5.6. UNet (For Segmentation)

While primarily used for medical image segmentation, UNet is also a backbone for identifying regions of interest (ROI) like lung nodules before classification. Many detection systems use UNet to first segment the lung area from CT/X-ray images, and then pass the segmented images through a classification CNN.

5.7. 3D CNNs

Since CT scans are volumetric, 3D CNNs are employed to capture spatial dependencies across multiple slices. Unlike 2D CNNs that process each image slice individually, 3D CNNs analyze the full scan as a 3D volume, leading to improved accuracy in detecting small and irregularly shaped nodules.

5.8. Hybrid Models (CNN + LSTM / CNN + Attention)

Hybrid models combine CNNs with LSTM (for sequence modeling) or attention mechanisms to focus on the most relevant image regions. These models improve classification performance by learning temporal or contextual dependencies in CT scan sequences.

doi: 10.48047/ijiee.2025.15.4.8

5.9. Transfer Learning Algorithms

Pre-trained CNNs such as InceptionV3, MobileNet, and EfficientNet have been used with transfer learning techniques for faster convergence and better performance on small lung cancer datasets. These models leverage features learned from large datasets like ImageNet and adapt them to medical imaging tasks.

5.10. Ensemble Models

To further improve performance and robustness, researchers have employed ensemble learning using multiple CNN architectures. Outputs from different models (e.g., ResNet + VGG + DenseNet) are combined using techniques like majority voting or softmax averaging to reduce model bias and variance.

6 RESULT:

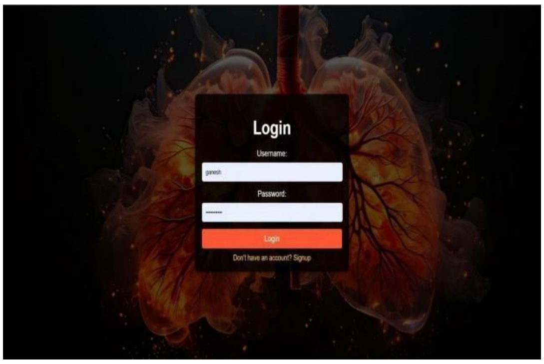

Fig 6.1 above result explain about user login to verify the lung cancer

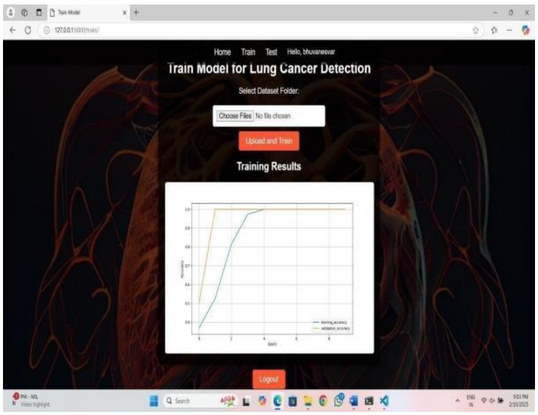

Fig 6.2 above result explain about training result to verify the lung cancer

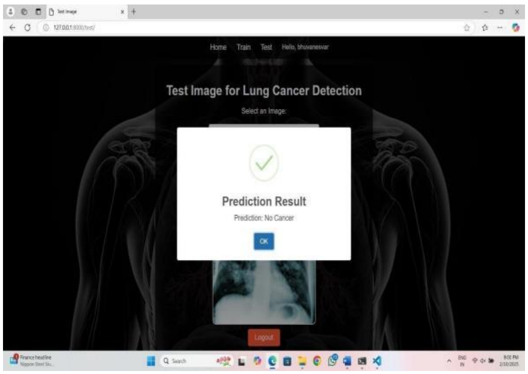

Fig 6.3 above result explain about predict the no lung cancer

7 CONCLUSION:

The lung cancer detection system using deep learning and Django provides a robust, efficient, and user-friendly solution for early detection of lung cancer. By integrating Convolutional Neural Networks (CNNs) with TensorFlow/Keras, the system leverages the power of deep learning to automatically analyze and classify medical images, such as chest X-rays and CT scans. The use of CNNs eliminates the need for manual feature extraction and improves the accuracy and speed of diagnosis. The web application not only simplifies the image upload and prediction process but also offers a platform for model training and performance monitoring. This system can potentially aid healthcare professionals in making faster, more accurate diagnoses, thus improving patient outcomes. With real-time prediction capabilities and the potential for future improvements in data quality and model robustness, this approach represents a significant step forward in the application of AI in medical.

8 REFERENCES

1. Hinton, G., Osindero, S., & Teh, Y. W. (2006). A fast-learning algorithm for deep belief nets. https://doi.org/10.1162/neco.2006.18.7.1527

2. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & van der Laak, J. A. W. M. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60-88. https://doi.org/10.1016/j.media.2017.07.005

3. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image

segmentation. Medical Image Computing and Computer-Assisted Intervention (MICCAI), 234-241. https://arxiv.org/abs/1505.04597

4. National Cancer Institute. (2023). Lung cancer statistics. Retrieved from https://www.cancer.gov/types/lung

5. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Ng, A. Y. (2017). CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning.

arXiv preprint arXiv:1711.05225. https://arxiv.org/abs/1711.05225

6. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

7. WHO – World Health Organization. (2023). Lung cancer burden and prevention strategies. Retrieved from https://www.who.int/news-room/fact-sheets/detail/lung-cancer

8. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444. https://doi.org/10.1038/nature14539

9. American Lung Association. (2023). Early detection and screening of lung cancer. Retrieved from https://www.lung.org/lung-health-diseases/lung-disease-lookup/lung- cancer

10. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition

(CVPR), 770-778. https://doi.org/10.1109/CVPR.2016.90

FIRST AUTHORS:

V.MANISHA pursuing her B.Tech in Computer Science And Engineering in Universal College Of Engineering And Technology.

M. HARIKA pursuing her B.Tech in Computer Science And Engineering in Universal College Of Engineering And Technology.

P. BHUVANESVAR pursuing his B.Tech in Computer Science And Engineering in Universal College Of Engineering And Technology.

B. SIVA SANKAR pursuing his B.Tech in Computer Science And Engineering in Universal College Of Engineering And Technology.

S. SRIKANTH REDDY pursuing his B.Tech in Computer Science And Engineering in Universal College Of Engineering And Technology.

SECOND AUTHOR:

J. SIVA PARVATHI M.Tech received her M.Tech degree and B.Tech degree in computer science and engineering. She is currently working as an Assist Professor in , Universal College Of Engineering And Technology.