International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 1 (2026)



ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 1, 2026
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research

AUTOMATED BLOOD CANCER CLASSIFICATION USING DEEP NEURAL NETWORKS

Dr.CHAITANYA KISHORE REDDY MADDIREDDY1, SANGU JAHNAVI2, DOKALA AJAY KUMAR3, GATTEM SAI4

#1 Professor of the Information Technology Department ,NRI Institute of Technology,Agiripalli

#2#3#4 B.Tech with Specialization of Information Technology in NRI Institute Of Technology,Agiripalli

Abstract: Improving survival rates for blood cancer, particularly leukemia, depends on early detection. Traditional diagnostic techniques depend on the labor-intensive and subjectively interpretable manual microscopic examination of peripheral blood smear pictures. Deep learning advances in recent years have made automatic detection more accurate and reliable. A labeled leukaemia dataset with four classes—Benign, Early, Pre (Pre-B), and Pro (Pro-B)—is used in this study to train and assess the model. Deep CNN designs like VGG16 and VGG19, which offer powerful feature extraction but are computationally demanding, were mostly used in earlier studies. MobileNetV2, a portable and effective deep learning model tailored for real-time diagnosis, is proposed in this work. For reliable classification, the suggested approach combines deep feature learning, image preprocessing, and augmentation. The results of the experiments show that the MobileNetV2 based model improves diagnosis speed while retaining excellent accuracy, which qualifies it for practical clinical implementation..

Index terms *- Blood Cancer Detection,Leukemia Classification,Deep Learning, Convolutional Neural Networks (CNN).*

Received: 31-01-2026 Accepted: 05-03-2026 Published: 12-03-2026

1. INTRODUCTION

When it comes to aquaculture production, China is the world leader, accounting for more than 70% of global production. 53.88 million tons of aquaculture were produced in China in 2021, making up 80.50% of all aquatic products [1]. The quality and production of aquatic commodities are directly impacted by the water quality of an aquatic environment because aquatic creatures carry out a variety of physiological functions in water, including respiration, eating, excretion, and reproduction. Since dissolved oxygen (DO) is necessary for aquatic creatures to survive and is utilized by their metabolism, it is a crucial indicator of water quality [3]. The healthy growth of farmed fish, shrimp, and other species can be impacted by excessive or insufficient DO, which can easily lead to disease outbreaks and even mass mortality, which would cause large financial losses for businesses [4-5]. Therefore, preventing water quality degradation, lowering aquaculture risk, and promoting the healthy and sustainable growth of intensive aquaculture all depend on anticipating dissolved oxygen concentrations and their trends, controlling dissolved oxygen concentrations promptly, and guaranteeing the healthy growth of aquatic products in a comfortable environment [6].

Numerous conclusions have been drawn from in-depth research on machine learning-based dissolved oxygen prediction models carried out by scholars both domestically and abroad. Liu, for instance, presented a forecasting model that combines the particle swarm optimization gravity search technique, empirical wavelet transforms, and the grey correlation degree. This model can more precisely interpret the dissolved oxygen trend, according to experiments [7]; Ren separated and denoised the original data using Variational Mode Decomposition (VMD) before supplying the decomposed data to a Deep Belief Network (DBN) for prediction [8]. Shi

proposed a dissolved oxygen prediction model based on clustering and an enhanced extreme learning technique [10], Huang proposed a hybrid model based on CEEMDAN-LZC and GOBLPSO to increase efficiency ,and Cao used principal component analysis (PCA) to filter the essential elements that would affect the dissolved oxygen. K-means clustering and GRU were then used to create the dissolved oxygen prediction model . A brand-new feature extraction-based approach was put out, improving predictive performance and yielding precise dissolved oxygen level predictions Support vector regression, multi-feature engineering, and optimization techniques are the foundation of Nong's proposed dissolved oxygen prediction model, and its application demonstrates that it can successfully raise prediction accuracy . Even though the dissolved oxygen prediction models listed above are capable of forecasting the dissolved oxygen content at future times, their slow computation speeds and inability to capture global contextual information make it difficult to meet the demands of precise and productive aquaculture production.

Complex biochemical reactions in the water column and intricate inter-factor modes of action are features of intensive aquaculture. By feeding all the variables into the dissolved oxygen prediction model, which results in a complex prediction model network structure and redundancy or overlapping information, the computational complexity and accuracy of dissolved oxygen prediction for water quality parameters in intensive aquaculture suffer. LightGBM is a tree-based boosting method that can greatly increase algorithm speed in big sample applications while also processing data efficiently and using less memory. For business financing risk prediction, Wang developed the LightGBM model, which may significantly increase risk prediction accuracy . In order to create a cryptocurrency price trend prediction model, Sun used LightGBM, which successfully assisted investors in making cryptocurrency investment decisions . For the short-term wind power forecasting model, Ren developed an attention mechanism based on CNN LSTM-LightGBM, which produces an accurate prediction .

2. LITERATURE SURVEY

1. Hypomethylation in Human Cancers

Early research on DNA methylation in cancer demonstrated that gene-specific hypomethylation distinguishes certain human cancers from their normal tissue counterparts. In this study, the authors examined primary tumor tissues and adjacent normal tissues without prior exposure to chemotherapy or radiation. Instead of measuring overall methylation levels, they focused on specific gene regions and found substantial hypomethylation in cancer cells in most patients studied. Additionally, metastatic samples showed progressive loss of methylation, suggesting that reduced DNA methylation may contribute to abnormal gene expression and tumor progression.

2. A Decade of Exploring the Cancer Epigenome

Baylin and Jones (2011) highlighted the central role of epigenetic mechanisms in cancer development, progression, and therapy. With the advancement of next-generation sequencing technologies, researchers gained deeper insight into alterations in the cancer epigenome. Their work emphasized that abnormalities in DNA methylation are often associated with mutations in genes regulating chromatin structure and metabolic processes. The authors also suggested that epigenetic modifications could serve as valuable biomarkers for cancer detection and prognosis, while epigenetic enzymes present promising therapeutic targets for reversing abnormal gene expression in tumors.

3. Tissue-Independent and Tissue-Specific DNA Methylation Patterns in Cancer

In a comprehensive genome-wide study, researchers analyzed DNA methylation patterns across multiple cancer types using large datasets. Their findings revealed that a significant proportion of aberrant methylation changes are shared across different cancers, indicating tissue-independent mechanisms. However, they also observed tissue-specific methylation alterations influenced by the chromatin state and histone modifications of the original normal tissue. The study concluded that pre-existing histone marks and promoter accessibility strongly predict the likelihood of abnormal methylation in cancer, highlighting the complex interaction between chromatin organization and epigenetic dysregulation.

4. Incomplete DNA Methylation and Transcriptional Memory in iPS Cells

Kim et al. (2010) investigated DNA methylation patterns in human induced pluripotent stem (iPS) cells derived from different somatic tissues. Their study showed that early-passage iPS cells retain a

transcriptional memory of their tissue of origin due to incomplete promoter methylation. This residual gene expression affects reprogramming efficiency and demonstrates that DNA methylation plays a critical role in maintaining or altering cellular identity. The authors emphasized that incomplete epigenetic reprogramming may have functional consequences in regenerative medicine applications.

5. Temporal and Regional Changes in DNA Methylation During Embryonic Development

Reik, Dean, and Walter (2001) examined dynamic changes in DNA methylation during mouse embryonic development. Their research demonstrated that early embryonic stages undergo global demethylation followed by progressive remethylation in embryonic and extraembryonic tissues. Germ cells were found to be significantly undermethylated, supporting the concept that epigenetic reprogramming is essential for developmental gene regulation. The study proposed that DNA methylation stabilizes gene expression patterns in somatic tissues while preserving developmental potential in germ line cells.

Collectively, these studies underscore the critical role of DNA methylation in cancer biology, cellular reprogramming, and embryonic development, emphasizing its importance in gene regulation and therapeutic research.

3. METHODOLOGY

i) Proposed Work:

An efficient, accurate, and real-time leukemia detection method based on MobileNetV2 is proposed to address the drawbacks of heavy CNN models. To reduce computational cost while keeping feature learning capability, MobileNetV2 leverages depth wise separable convolutions. The model undergoes thorough preparation procedures, including noise removal, normalization, cell segmentation, and augmentation, to enhance generalizability, while being trained on the identical four-class leukaemia dataset (Benign, Early, Pre, and Pro). The suggested system is well-suited for cloud platforms, mobile-based diagnostic systems, and clinical assistance tools because to its reduced memory usage, high classification efficiency, and quicker training time compared to VGG16 and VGG19. Without sacrificing diagnostic accuracy, the method seeks to aid pathologists by providing fast, automated, and dependable predictions..

ii) System Architecture:

The proposed architecture for Leukemia Detection using MobileNetV2 follows a systematic deep learning pipeline designed for accurate classification of blood smear images. The process begins with the input of microscopic blood smear images, which serve as the primary data source for detecting leukemia. These images first undergo preprocessing, where they are resized to a fixed dimension, normalized to standardize pixel intensity values, and enhanced to reduce noise and improve visual clarity. This step ensures that the images are in a consistent and suitable format for model training.

After preprocessing, data augmentation techniques such as rotation, flipping, zooming, and shifting are applied to artificially expand the dataset and improve model generalization. Augmentation helps prevent overfitting and enhances the model's ability to perform well on unseen data.

The processed images are then fed into the MobileNetV2 architecture, which serves as the core feature extraction and classification model. MobileNetV2 consists of convolutional layers that extract spatial features from the images, followed by bottleneck layers that use depthwise separable convolutions to reduce computational complexity while maintaining high performance. These bottleneck layers improve efficiency and make the model suitable for real-time and resource-constrained environments. The extracted deep features are passed to fully connected layers, which perform the final classification.

Finally, the system outputs the leukemia detection result, identifying whether the blood smear image corresponds to a specific leukemia class or a normal category. Overall, this architecture provides a lightweight, efficient, and accurate deep learning framework for automated leukemia detection suitable for clinical applications.

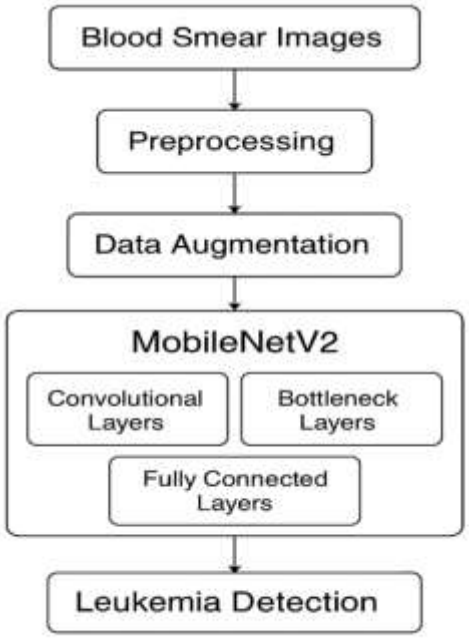


Fig.1. Proposed Architecture

iii) MODULES:

1. Blood Smear Image Acquisition Module

This module is responsible for collecting and managing microscopic blood smear images used for leukemia detection. The images are obtained from publicly available medical datasets or laboratory sources. Proper labeling of images into different classes such as Benign, Early, Pre-B, and Pro-B is ensured to maintain dataset accuracy. The collected images are organized systematically for training, validation, and testing purposes.

2. Preprocessing Module

The preprocessing module prepares raw blood smear images for model training. It includes resizing images to a fixed resolution (e.g., 224×224), normalizing pixel values to improve convergence, and performing noise reduction or contrast enhancement to improve image clarity. This module ensures uniform input format and enhances important visual features required for accurate classification.

3. Data Augmentation Module

This module increases dataset diversity by applying transformations such as rotation, flipping, zooming, shifting, and brightness adjustments. Data augmentation helps reduce overfitting and improves the model's generalization ability by exposing it to varied image patterns during training.

4. Feature Extraction Module (MobileNetV2)

The core module of the system utilizes the MobileNetV2 architecture for efficient feature extraction. It consists of convolutional layers that capture spatial features from images and bottleneck layers that use depthwise separable convolutions to reduce computational cost. This lightweight structure ensures faster processing while maintaining high classification accuracy, making it suitable for real-time applications.

5. Classification Module

The classification module includes fully connected layers that interpret the extracted features and assign the image to one of the predefined leukemia classes. A Softmax activation function is used to generate probability scores for each class, and the class with the highest probability is selected as the final prediction.

6. Leukemia Detection Output Module

This final module displays the prediction result along with confidence scores. The output can be integrated into a clinical decision-support system, enabling healthcare professionals to make timely and informed diagnostic decisions.

iv) ALGORITHMS:

1. Image Preprocessing Algorithm

The preprocessing algorithm prepares raw blood smear images for deep learning analysis. First, each image is resized to a fixed dimension (e.g., 224 × 224 pixels) to ensure uniform input size. Next, pixel intensity values are normalized to a standard range (such as 0–1) to improve numerical stability during training. Noise reduction and contrast enhancement techniques may also be applied to highlight important cellular structures. The output of this algorithm is a clean, standardized image suitable for model input.

2. Data Augmentation Algorithm

The data augmentation algorithm artificially increases dataset size and diversity. It applies transformations such as rotation, horizontal/vertical flipping, zooming, shifting, and brightness adjustments to original images. These transformations generate multiple variations of the same image while preserving its class label. This helps reduce overfitting and improves the model's generalization capability.

3. MobileNetV2 Feature Extraction Algorithm

MobileNetV2 uses depthwise separable convolutions to efficiently extract features from images. The algorithm first applies standard convolution to detect basic patterns like edges and textures. Then, depthwise convolution filters each input channel separately, followed by pointwise convolution to combine features across channels. Bottleneck layers with inverted residual connections are used to reduce computational cost while maintaining representational power. This structure enables efficient feature extraction suitable for lightweight and real-time applications.

4. Forward Propagation and Classification Algorithm

During forward propagation, the preprocessed image passes through convolutional and bottleneck layers to extract deep features. These features are flattened and fed into fully connected layers. A Softmax activation function is applied at the output layer to compute probability scores for each leukemia class (Benign, Early, Pre-B, Pro-B). The class with the highest probability is selected as the final prediction.

5. Training Algorithm (Backpropagation with Adam Optimizer)

The training process uses backpropagation to update model weights based on prediction errors. A loss function such as categorical cross-entropy calculates the difference between predicted and actual labels. The Adam optimizer adjusts learning rates adaptively for faster convergence. This iterative process continues for multiple epochs until the model achieves optimal performance on training and validation data.

6. Model Evaluation Algorithm

After training, the evaluation algorithm measures model performance using metrics such as accuracy, precision, recall, and F1-score. Predictions on the test dataset are compared with true labels to compute these metrics. This step ensures that the model generalizes well to unseen data and performs reliably in real-world clinical scenarios.

4. EXPERIMENTAL RESULTS

Accuracy: The ability of a test to differentiate between healthy and sick instances is a measure of its accuracy. Find the proportion of analysed cases with

true positives and true negatives to get a sense of the test's accuracy. Based on the calculations:

Accuracy = TP + TN /(TP + TN + FP + FN)

$$Accuracy = \frac{(TN + TP)}{T}$$

Precision: The accuracy rate of a classification or number of positive cases is known as precision. Accuracy is determined by applying the following formula:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\Pr e\,cision = \frac{TP}{(TP + FP)}$$

Recall: The recall of a model is a measure of its capacity to identify all occurrences of a relevant machine learning class. A model's ability to detect class instances is shown by the ratio of correctly predicted positive observations to the total number of positives.

$$Recall = \frac{TP}{(FN + TP)}$$

F1-Score: A high F1 score indicates that a machine learning model is accurate. Improving model accuracy by integrating recall and precision. How often a model gets a dataset prediction right is measured by the accuracy statistic..

$$F1 = 2 \cdot \frac{(Recall \cdot \Pr e\,cision)}{(Recall + \Pr e\,cision)}$$

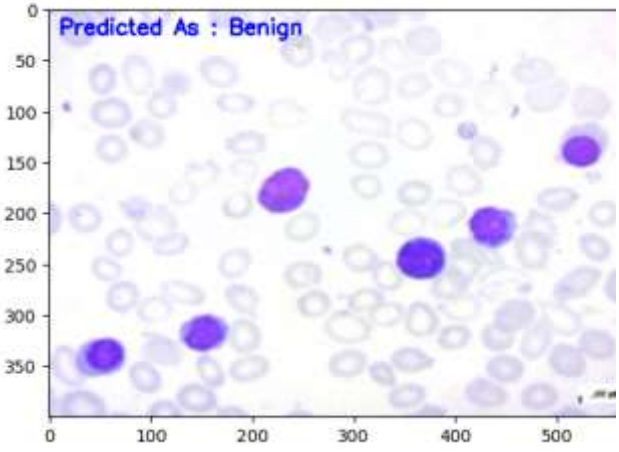


Fig 2:Disease Detected

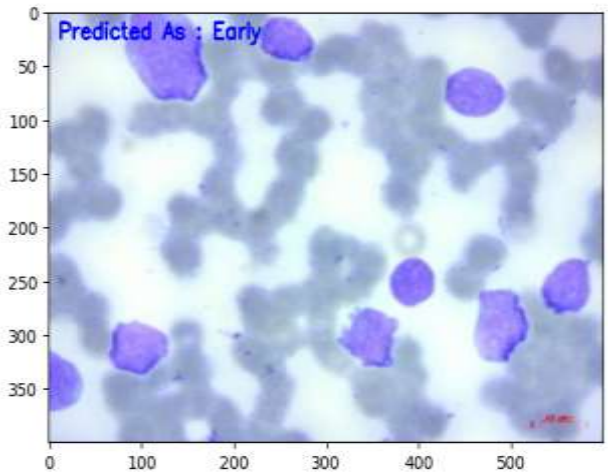


Fig 3:Disease Detected

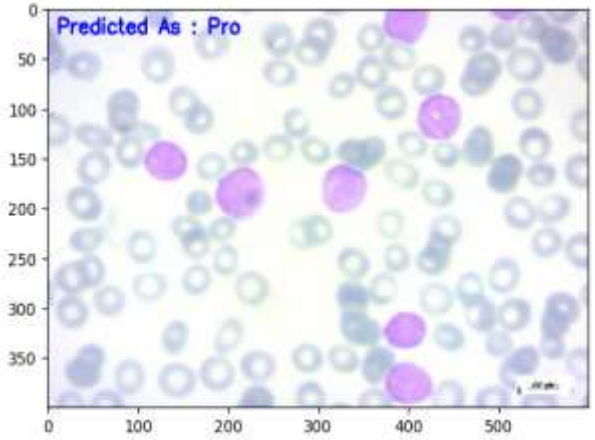


Fig 4:Disease Detected

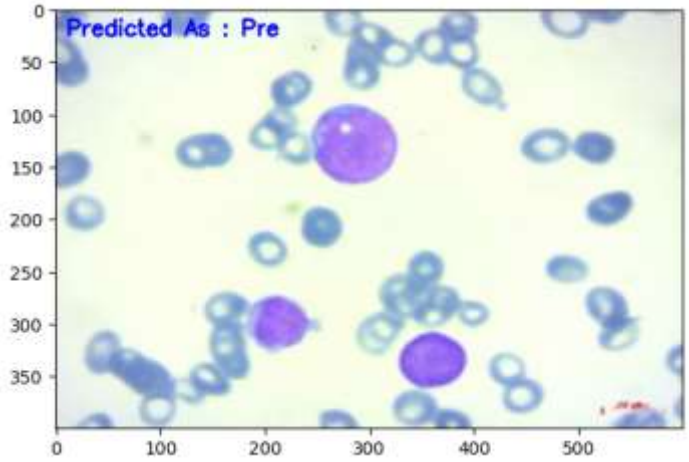


Fig 5:Disease Detected

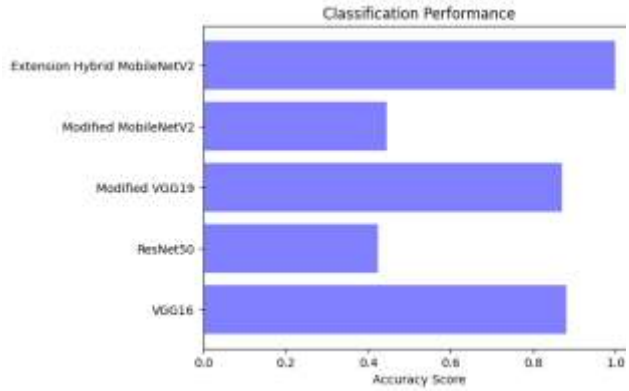


Fig 6:Accuracy Comaprision

5. CONCLUSION

Automated leukaemia identification utilizing a four-class dataset covering benign, early, pre-, and pro malignancy stages is presented in this paper as a deep learning method. Despite the robust feature extraction demonstrated by previous algorithms using VGG16 and VGG19, their processing weight limits its usability in real-time. In response to these issues, we provide a MobileNetV2-based model that is ideal for widespread clinical use because it is fast, accurate, and lightweight. The results show that MobileNetV2 is the best option for healthcare facilities with limited resources since it combines efficiency with performance. In order to aid pathologists, speed up screening, and improve early identification of leukaemia, this research contributes to the development of AI-assisted diagnostic tools..

6. FUTURE SCOPE

The proposed MobileNetV2-based leukemia detection system can be further enhanced in several directions to improve clinical applicability and robustness. Future work can focus on expanding the dataset by including larger and more diverse blood smear images collected from multiple healthcare institutions to improve model generalization across different populations and imaging conditions. Integration of multi-modal data, such as patient clinical history, laboratory reports, and genetic biomarkers, could further enhance diagnostic accuracy and provide comprehensive decision support. Implementing explainable AI (XAI) techniques, such as Grad-CAM visualization, would help pathologists understand the regions influencing model predictions, thereby increasing trust and transparency. The system can also be optimized for deployment on mobile or edge devices to enable real-time screening in rural or resource-limited healthcare settings. Additionally, incorporating automated severity grading and treatment recommendation modules could transform the system into a complete clinical decision-support platform for early leukemia diagnosis and management.

REFERENCES

1. Feinberg AP, Vogelstein B, et al. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature. 1983;301(5895):89–92. doi: 10.1038/301089a0 - DOI - PubMed

2. Baylin SB, Jones PA. A decade of exploring the cancer epigenome—biological and translational implications. Nature Reviews Cancer. 2011;11(10):726–734. doi: 10.1038/nrc3130 - DOI - PMC - PubMed

3. Chen Y, Breeze CE, Zhen S, Beck S, Teschendorff AE. Tissue-independent and tissue-specific patterns of DNA methylation alteration in cancer. Epigenetics & chromatin. 2016;9(1):1. - PMC - PubMed

4. Marella, V. C., Veluru, S. R., & Erukude, S. T. (2025, September). FedOnco-Bench: A Reproducible Benchmark for Privacy-Aware Federated Tumor Segmentation with Synthetic CT Data. In 2025 4th International Conference on Innovative Mechanisms for Industry Applications (ICIMIA) (pp. 870-876). IEEE.

5. Ohi Y, Qin H, Hong C, Blouin L, Polo JM, Guo T, et al. Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. Nature cell biology. 2011;13(5):541–549. doi: 10.1038/ncb2239 - DOI - PMC - PubMed

6. Mallick, P. (2025). AgentAssistX: An Agentic Generative AI Framework for Real-Time Life & LTC Insurance Advisory, Risk

Scoring, and Compliance Validation in Cloud-Native Environments.
7. Monk M, Boubelik M, Lehnert S. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. Development. 1987;99(3):371–382. - PubMed

Author profile



Dr.Chaitanya Kishore Reddy Maddireddy is currently working as a Professor in the Department of Information Technology at NRI Institute of Technology(NRIIT), Agiripalli, Vijayawada, Andhra Pradesh, India. Has 20 years of academic experience and 2 years of industry experience. He earned master's degree in Computer Science and Engineering and a Doctorate in Computer Science and Engineering. He has Published more than 100 research papers in various National,International Journals and International Conferences, authored 5 textbooks. He is a member in ISTE, CSI, and IAENG. His research areas are Mobile Ad-hoc Networks, Cloud Computing, Artificial Intelligence and Machine Learning. He has made significant contributions to the field through his Research, Teaching and Publications.
Email-chkishore.0007@gmail.com



SANGU JAHNAVI REDDY is a B.Tech student specializing in Information Technology at NRI Institute of Technology. She has strong knowledge in Python and SQL, with hands-on experience in Flask, Tkinter, MySQL,and XAMPP for developing secure applications. She completed a Python Internship at NIELIT, Tirupati, contributing to the development of "MedLink – Secure Hospital Management System" using Python, Flask, and MySQL. She also completed a Virtual Internship with VOIS & Vodafone Idea Foundation, gaining knowledge in data analytics and machine learning using Python. She is actively involved in leadership roles such as Treasurer of Helping Hands Club, Event Coordinator at the college, and Class Representative. She is passionate about learning new technologies and contributing effectively in a professional environment. She demonstrates strong problem-solving abilities and a keen interest in backend development, with a commitment to continuous improvement and building practical solutions to real-world challenges.



DOKALA AJAY KUMAR is a B.Tech student with Specialization in Information Technology at NRI Institute of Technology. With a strong interest in emerging technologies, he has developed skills in Python, Web Development, and Cloud Computing. he holds an AWS certification, strengthening her knowledge in cloud platforms and deployment practices. As part of his academic curriculum, he completed a mini project titled "Plant Growth Monitoring System using AI & ML," where he applied Artificial Intelligence and Machine Learning techniques to monitor and analyze plant growth efficiently. he is passionate about learning new technologies and continuously improving her programming and problem-solving skills.



GATTEM SAI B.Tech Student with specialization of information Technology in NRI Institute of Technology .I have a strong foundation in Python programming, data structures, object-oriented programming, and database management systems.I have worked on projects involving data analytics and machine learning, such as a doctor visit recommendation system using data analytics and AI, a weather tracking system, and an income prediction model using machine learning. Through these projects, I gained hands-on experience with Python libraries like NumPy and Pandas and improved my problem-solving and analytical skills.I have also completed certifications in Google Data Analytics, Google IT Support, Python for Data Science, and Cloud Computing, which helped me understand both software development and IT fundamentals. I am a quick learner, enjoy working in teams, and I am eager to start my career as an entry-level software engineer where I can apply my skills and continue learning.