



**NARASARAOPETA ENGINEERING COLLEGE (AUTONOMOUS)**  
**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

2021-2022

<b>Batch Number</b>	DB-13
<b>Team Members</b>	Sk. Shajahan(23475A0519) J. Surya Teja(22471A05M6) N.V.C. Sudheer(23475A0522)
<b>Guide</b>	Ch. Rajani (MCA, MTech)
<b>Title</b>	IMAGE CAPTIONING
<b>Domain/Technology</b>	DEEP LEARNING
<b>Base Paper Link</b>	<a href="https://ieeexplore.ieee.org/document/10555151/figures#figures">https://ieeexplore.ieee.org/document/10555151/figures#figures</a>
<b>Dataset Link</b>	<a href="https://www.kaggle.com/code/mpwolke/rsicd-image-captioning/input">https://www.kaggle.com/code/mpwolke/rsicd-image-captioning/input</a>
<b>Software Requirements</b>	Browser: Any latest browser like Chrome Operating System: Windows 7 Server or later Python (COLAB)
<b>Hardware Requirements</b>	SystemType: Intel Core i5 or above RAM: 8 GB Number of cores:5 Number of Threads: four
<b>Abstract</b>	<p>Image captioning combines computer vision and NLP to generate accurate textual descriptions of images. For remote sensing, this aids in environmental monitoring and urban planning. Traditional single-model approaches face challenges in generalization, prompting the use of ensemble deep learning methods for robust captioning.</p> <p>This project implements an NLP-based fusion approach using the RSICD dataset (10,921 satellite images with 5 captions each). By integrating ViT, ResNet, and GPT-2 models with BERT, CLIP, and VAE fusion strategies, we achieve a CIDEr score of 0.86 (86%), outperforming single-model methods. The system, built with PyTorch and Hugging Face, ensures scalable and efficient remote sensing applications.</p>

Signature of the student(s)

Signature of the Guide

Signature of the project coordinator