RESEARCH PAPER

Attention Enhanced InceptionNeXt-Based Hybrid Deep Learning Model for Lung Cancer Detection

[1]V. Naga Saranya, [2]Dr. P. Venkata Hari Prasad

[1] M.Tech CSE, student, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur - 522302.

vns2099@gmail.com

[2] Associate Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur - 522302

pvhariprasad@kluniversity.in

Abstract— Lung cancer is the leading cause of cancer death which demands timely and accurate diagnosis to improve patient outcomes. Computed tomography (CT) scans are often poorly localized and identified using manual interpretation and standard deep learning methods to find multi-scale features. Two publicly available datasets, which are the IQ-OTH/NCCD Lung Cancer Dataset and the Chest CT-Scan Images Dataset are used to perform experiments. It is proposed to use the hybrid attention-enhanced CNNViT framework based on ResNet50, DenseNet169, EfficientNetV2-Medium, ConvNeXt-Base, InceptionNeXt-Base, MobileViT-Small, ConViT-Base, Swin-Base, MaxViT-Base, and DeiT3-Base to classify images, and YOLOv5, YO Preprocessing includes image resizing to 299x299, image data augmentation, normalizing tensors and stratified data division, at the same time YOLO datasets are cleaned with bounding box labels. GradCAM generates heatmaps highlighting noteworthy regions, whereas a Flask-based interface allows a thorough interaction with the user. ConvNeXt-Base has a peak classification rate of 99.09% on the IQ-OTH/NCCD data, and InceptionNeXt-Base has a peak accuracy rate of 99.01% on the chest CT-scan dataset. The highest mean average precision (mAP) of 72.8% is achieved with YOLOv5 in detection. The approach exhibits enhanced robustness, equitable performance, and elucidated predictions by the integration of categorization and detection inside a cohesive system.

“Keywords— Lung Cancer Detection, Computed Tomography, Hybrid CNN–Vision Transformer, Attention Mechanisms, Deep Learning, Explainable AI”.

HOW TO CITE THIS ARTICLE: NAGA SARANYA V, VENKATA HARI PRASAD P. ATTENTION ENHANCED INCEPTIONNEXT-BASED HYBRID DEEP LEARNING MODEL FOR LUNG CANCER DETECTION. INT J DRUG DELIV TECHNOL. 2026;16(50S): 69-75. DOI: 10.25258/IJDDT.16.50S.10

I. INTRODUCTION

Lung cancer remains among the most prevalent and deadly cancers worldwide, with a huge effect on mortality rates. The disease is caused by unusual cell growth in lung tissues and often progresses rapidly to cause severe breathing complications. In spite of the improvements in the medical field, lung cancer continues to affect millions of people every year, and the survival rates largely depend on the timely diagnosis and timely treatment. The epidemiological evidence shows that there is an increasing number of diagnosed cases and associated deaths, which is why there is a need to develop improved diagnostic methods that will improve clinical decision-making and patient care [1]. These subdivisions of lung cancer into primary subgroups highlight the complexity of accurate diagnosis and development of the treatment strategy [2].

Conventional diagnostic techniques, such as screening through imaging and invasive clinical interventions, play a crucial role in identifying the lung issues. However, because of the methods, they often involve significant costs, long analysis times, and potential suffering of patients. In addition, small nodules in the lungs or small pulmonary cancers are sometimes not detected or misunderstood due to their unobtrusive appearance. Human judgment of imaging data is based on professional skill and prone to inconsistency and error, and this limits the reliability of a diagnosis. Although advancement in technology has improved imaging and produced better diagnostic systems, the need to have automated and accurate diagnostic assistance systems has remained a major hitch in the medical imaging analysis [3], [4], [5].

In the recent past, development of advanced computational techniques has enabled automatic interpretation of medical imaging data, which has brought about tremendous improvement in disease diagnosis and classification accuracy. High level learning models have demonstrated the ability to identify complex visual characteristics and identify tiny defects which can be overlooked by human evaluation. Such methodologies have drawn increased attention due to their ability to enhance diagnostic efficacy and consistency across different imaging modalities. The primary objective is to develop an effective and powerful diagnostic system capable of adequately identifying the lung abnormalities in medical imaging, and surpass the constraints of existing automated systems [6], [7], [8].

The use of advanced computer analysis in the diagnosis of lung cancer has significant therapeutic and social implications. Proper and prompt diagnosis can enable prompt treatment, higher of survival, and reduced cost of treating illnesses at an advanced stage. Moreover, an increased diagnostic reliability would assist physicians to select the most optimal treatment to apply on an individual patient and minimized the scope of views, which are involved in clinical assessment. The introduction of high-image analysis systems makes it easier to launch scalable screening programs and access to diagnostic services, particularly in regions with limited medical expertise. These inventions have the potential of transforming the diagnostic processes and improving patient outcomes across the globe [9], [10].

II. LITERATURE REVIEW

The recent advances in artificial intelligence have drastically influenced automated lung cancer detection through medical imaging. In [11], it is clearly indicated that deep learning algorithms are rapidly being employed to enhance accuracy and reduce the human factor in lung cancer diagnosis. The survey of [12] explores various deep learning architectures applied to CT imaging, demonstrating the improved feature extraction and classification performance compared to traditional machine learning tools. Although these studies provide much important information about the existing methods, they mainly focus on the summarization of structures rather than addressing the problems related to multi-scale feature representation and lesion localization.

Many studies have introduced tailored attention-based and neural network models to enhance the effectiveness of diagnosis. MorphAttnNet framework, proposed in [13], integrates morphological feature extraction with attention mechanisms to help improve the accuracy of subtype classification, and showed promising results when detecting structural tumor features. Another approach in [14] introduces a two-way recurrent neural network enhanced with a bio-inspired algorithm to enhance the classification stability and accuracy. GoogLeNet-AL framework in [15] is an adaptive automated detection framework that improves adaptation of the model to various situations in imaging. These methods have good classification abilities, but often rely on single-model designs, which limits their ability to extrapolate to different datasets and imaging complexities.

Ensemble learning techniques have also been of interest in improving the detection strength and reliability of classification. The ensemble-based deep learning system described in [16] incorporates multiple neural configurations to complement diagnostic accuracy of thoracic CT images, and exhibits superior performance through model diversity. The interpretable architecture of [17] introduces explainable artificial intelligence to improve transparency in diagnostic predictions, which contributes to a better clinical trust and interpretability. At the same time, federated learning coupled with blockchain technology in [18] enhances data privacy and efficiency of distributed training, reducing the privacy concerns of medical data sharing. Despite such progress, explainable and ensemble frameworks often increase the computational load and can require significant amounts of computer resources, therefore limiting real-time clinical application.

Advanced studies have been done on new network architecture to learn more features and classify them. The capsule-network-based ensemble model introduced in [19] extends the spatial feature representation and improves detection based on the hierarchical linkage of imaging data. Lung-EffNet, introduced in [20], makes use of the EfficientNet architecture to achieve high classification rates and to achieve computing economy. Although these techniques perform highly, there are still challenges in achieving a trade-off between computing economy, multi-scale feature learning and accurate localization of lesions in various datasets.

In spite of the significant progress in automated lung cancer detection, the existing methodologies often face the challenge of concurrently measuring local and global image features at the same time without compromising computing efficiency and model interpretability. Many of these frameworks focus on categorization, but do not provide comprehensive lesion detection and visualization capabilities. This question addresses these issues by offering a combined hybrid learning system that will improve the extraction of multiple features, improve reliability in detection and provide diagnostic information that can be explained. This approach aims to build resilience and offer balanced performance in a variety of CT imaging datasets thereby enabling reliable clinical decisions.

III. MATERIALS AND METHODS

The system will aim to provide accurate classification of lung cancer and anomalies detection using IQ-OTH/NCCD Lung Cancer Dataset and Chest CT-Scan Images Dataset. The framework processes CT images through a structured pipeline comprising data ingestion, quality improvement, transformation and equitable partitioning of data sets to enable reliable model training and validation. Different baseline deep learning frameworks are evaluated to establish performance rates, followed by a simplified hybrid CNNVision Transformer model with InceptionNeXt blocks added and combined grid and block attention to augment multi-scale feature extraction and contextual representation. The framework improves diagnostic potential by combining the YOLO-based object detection models to the automated localization of suspicious regions and GradCAM-based explainable visualization that highlights important features which influence classification decisions. A Flask-based interface enables straightforward set-up and interaction with users ensuring accessibility and scalability. The hybrid system enhances diagnostic accuracy, interpretability and stability, and is a comprehensive and clinically useful model of automated lung cancer analysis on a wide range of CT image data.

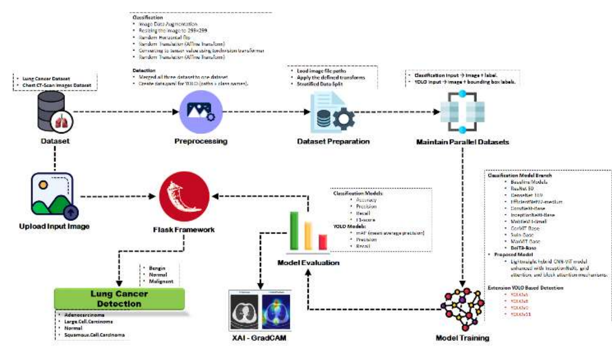


Fig. 1. Proposed Architecture

The system architecture includes a hybrid deep learning architecture that is aimed at lung cancer classification and anomaly detection. The input CT image is sequentially processed by setting up feature extension at different stages that encompass hybrid blocks which contain convolutional and attention-based processing. Inception depthwise convolution obtains multi-scale spatial representation, and the grid and block attention modules enhance the contextual representation and focus on salient areas. The hierarchical feature refinement is implemented through iterative steps of hybridization, followed by pooling and completely connected layers. The end output layer will classify photos into a range of lung cancer types, which ensures effective and accurate diagnosis.

A) Dataset Collection

The experimental paper uses the IQ-OTH/NCCD Lung Cancer Dataset and the Chest CT-Scan Images Dataset that was obtained in the context of the publicly available medical imaging repositories. The IQ-OTH/NCCD data has 1,190 CT scan slices of 110 patients who were categorized into normal, benign and malignant groups and collected using standardized clinical imaging methods. The Chest CT dataset includes the images of adenocarcinoma, large cell carcinoma, squamous cell carcinoma, and normal cases in the JPG/PNG format, as well as pre-defined training, validation, and testing sets. The combined datasets provide diverse imaging characteristics, which support reliable training and testing of the model in a range of lung cancer types.

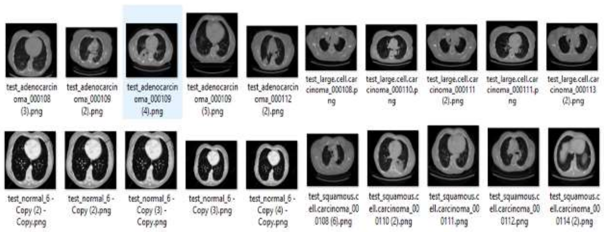


Fig.2 IQ-OTH/NCCD - Lung Cancer Dataset

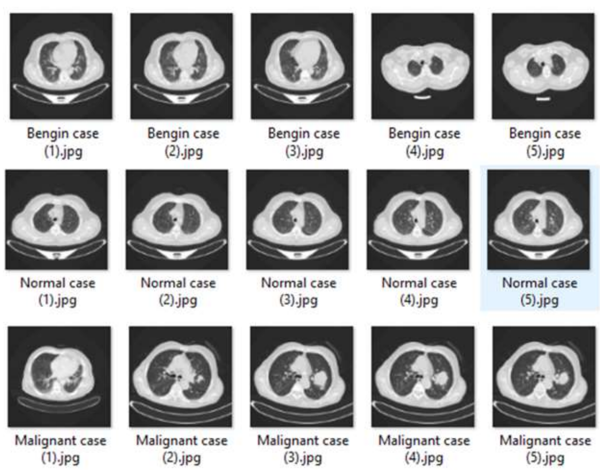
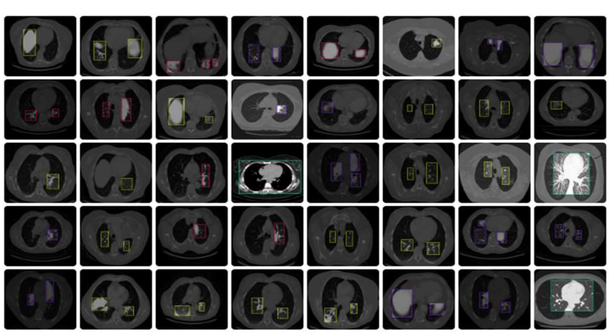


Fig.3 Chest CT-Scan images Dataset



B) Pre-Processing

Enhancing the quality of data, generalizing the models, and ensuring dependable learning depends on effective preprocessing. This on-the-fly pipeline ensures that the imaging data is more predictable, balances the distribution of data in datasets, and provides structured data to applications that consume the classification and detection.

Image Data Augmentation: To ensure that datasets are more varied and to enable models to operate under more situations, image data enhancement and standardization is performed. In addition, all the input images are downsampled to the same resolution to ensure that the set appears the same. The changes in picture location and orientation are made to appear as in real life by extra enhancement techniques such as horizontal flipping and spatial translation. Then the pictures are converted into quantified numerical representations which can be computed with. These alterations enable the model to be less prone to overfitting, more capable of adapting to changes in space, and more capable of learning various visual patterns.

Detection Dataset Integration and Configuration: In the case of the detection job, multiple picture datasets are grouped into a single one to make the sample more diverse and present a greater number of lung disorders. The collection of consolidations is organized, with definitions of annotations that associate the image locations with the class labels. A common dataset structure is defined to outline data paths and category definitions. This phase will ensure that it is compatible with object identification systems, enhances training efficiency and increases the ability of the model to identify and localize suspicious areas with a high level of reliability and across various imaging sources.

Dataset Loading: A systematic way of loading the datasets and transforming them is implemented to organize and prepare imaging data to be used in training models. Here you will obtain the whereabouts of the picture files and do some adjustments to ensure that the data style and quality are identical. The transformation chain ensures that all the samples have been preprocessed in such a manner that there is less variation and enhances training stability. Automation of data preparation, in the process of model creation, accelerates the process of processing data, facilitates its reuse, and ensures the integrity of the dataset structure.

Stratified Data Partitioning: Stratification of partition data is possible, that is, the data is divided into training and evaluation sets with the same number of classes in each set. This process ensures every subgroup is representative of the characteristics of the whole dataset, therefore, preventing the issue of class imbalance that may negatively affect training a model. Balanced partitioning enhances the reliability of models, and assessment fairness, as well as making it easier to generalize to unobserved data sets. Representative distributions in subsets would enable the consistency in performance and alleviate bias in model validation and testing.

Parallel Dataset Structuring for Classification: The structure of classification tasks is set up in parallel datasets, with a diagnostic label linked to each picture that is processed. This organized way of thinking will make it easy to get data and make sure that pictures that are inputted are correctly matched with categories that are outputted. Its parallel structure allows the training workflows to be easily connected to classification models and simplifies their management. The phase facilitates data access, enhances accuracy in training, and ensures a consistent performance of different classification architectures through constant input-label associations.

C) Algorithms

Classification:

ResNet-50: ResNet-50 simplifies the process of identifying things and this is achieved by taking advantage of the residual connections to maintain the learning of deep networks stable and retain the hierarchical representations of features. It works well to extract visual patterns since it maintains gradient flow between layers. This makes generalization better and makes sure that complex structure differences in imaging data are correctly recognized.

DenseNet-169: DenseNet-169 further improves learning features with dense connectivity that provides features the ability to be reused effectively and enhances the spread of gradients. The design enhances stability in classifications and efficiency in the parameters and is able to collect detailed and abstract image representations, thus increasing the discriminative power and reducing overfitting.

EfficientNetV2-Medium: EfficientNetV2-Medium improves the performance of the classification by optimizing the depth, width, and resolution of the network. It supports effective feature extraction with reduced computing complexity without losing contextual and spatial information thus encouraging faster training convergence and reliable generalization of diverse patterns of images.

ConvNeXt-Base: ConvNeXt-Base is a more efficient convolutional learning model that incorporates some design aspects of transformers. It gives a better picture of spatial traits and a better sense of the situation, which makes classification stronger and more consistent in complex visual systems with the best architectural designs.

InceptionNeXt-Base: InceptionNeXt-Base enhances multi-scale feature extraction, using parallel processing channels to combine multiple receptive fields. This architecture is able to store successfully local detail and global contextual information. This enhances classification accuracy and flexibility in imaging characteristics that have intricate structure patterns and certain imaging characteristics.

MobileViT-Small: MobileViT-Small is a hybrid of light convolutional operations with attention-based transformer to maximize computing efficiency and contextual learning. It models long-range dependencies with spatial inductive biases, and can be used to scale with reliable classification accuracy in resource-constrained environments.

ConViT-Base: ConViT-Base is an innovative feature learning model that combines convolutional inductive biases with transformer-based attention. It dynamically balances local spatial features with global contextual features, thereby enhancing convergence stability, interpretability, and classification reliability of various visual patterns.

Swin-Base: Swin-Base makes use of the hierarchical transformer processing and shifted window attention to achieve the successful capture of multi-scale spatial interdependence. It enhances feature aggregate and contextualization besides reducing the complexity of computations, and therefore, boosts the accuracy of classification and robust generalization of high-resolution visuals.

MaxViT-Base: MaxViT-Base combines the convolutional procedures with dual-axis attention mechanisms to simultaneously gather local attributes and global information. The representation by this combined form facilitates easier display of features, facilitates classification and allows description of complex visual interactions in a large number of image structures.

DeiT3-Base: According to the data-efficient transformer learning and attention optimization, DeiT3-Base enhances the classification performance. It is concerned with obtaining unique global characteristics and maintaining the training constant, which allows the training to produce more stable classification and enhanced generalization to a broader set of visual recognition tasks.

Lightweight Hybrid CNN–ViT with Attention: It takes advanced attention strategies and mixes the extraction of convolutional features with transformer-based contextual modeling to make the CNNViT lightweight hybrid model. It does a good job of showing multi-scale spatial information and global relationships. It also enhances classification, makes it more understandable and scalable without compromising efficiency of the parameters.

Detection:

YOLOv5: YOLOv5 is an object detector which is capable of identifying the position of an item and classify it in real time in a single system. It enhances the rapidity and precision of reasoning as it learns spatial characteristics more rapidly. This makes it possible to reliably find and locate problems in a variety of visual settings.

YOLOv8: YOLOv8 is superior in locating things since it does it without the use of anchors and has superior means of assembling features. It not only allows finding items of various sizes and shapes with high accuracy, but it also increases generalization without reducing the speed of inference or detection accuracy.

YOLOv9: YOLOv9 is designed to achieve a more stable detection by simplifying the way one can learn to encode features and enhance gradient flow between network levels. It not only lets you accurately locate and group complex visual patterns, but it also makes sure that the accuracy and memory of detections are the same in all situations.

YOLOv11: YOLOv11 is an improvement on single-stage detection because it simplifies the extraction of features and predictions. It enhances localization accuracy and computing efficiency, allowing one to locate objects more reliably and scale, and yet still be able to perform complex visual tasks in real time.

D) Integration of XAI & Flask:

The incorporation of Explainable Artificial Intelligence (XAI) improves model transparency by offering visual elucidations of prediction results. The visualization of the model is done using GradCAM to generate heatmaps that highlight important parts of the picture that influence the result of classification and detection. This interpretability process helps in understanding model reasoning by identifying discriminative factors that help to identify abnormalities. This visualization improves clinical reliability by enabling confirmation of areas of model attention, therefore, promoting trust and facilitating the confirmation of automated diagnostic findings based on intuitive visualization.

The Flask structure is also added in order to provide a simple deployment interface enabling easy communication with trained models. It supports the entry of picture inputs, processing of predictions, and displaying results in a unified environment. This streamlined online interface improves accessibility, enabling users to do real-time analysis without the need for specialist technical skills. The framework of deployment ensures scalability, integration of models, and fluid communication between the prediction models in the back end and the visualization elements in the front end.

IV. Experimental Results

Accuracy: The accuracy of a test is its ability to discriminate patient and healthy cases correctly. In determining the accuracy of a test, the ratio of the true positives and true negatives of the evaluated cases must be computed. This can be mathematically stated as:

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \quad (1)$$

Precision: Precision evaluates the ratio of correctly identified cases of the identified positive cases. The equation of calculating precision therefore is as follows:

$$Precision = \frac{\text{True Positive}}{\text{True Positive } + \text{False Positive}} \quad (2)$$

Recall: The recall is a machine learning metric that evaluates how well a model can identify all relevant examples of a certain category. It is the percentage of correct value positives predicted to the total actual positives, which provide information on the effectiveness of the model to detect the presence of a particular type of class.

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad (3)$$

F1-Score: F1 score is a statistical measure that is used to assess the accuracy of a machine learning model. It combines the accuracy and recall measures of a model. The accuracy measure is used to measure how often a model makes accurate predictions over the entire dataset.

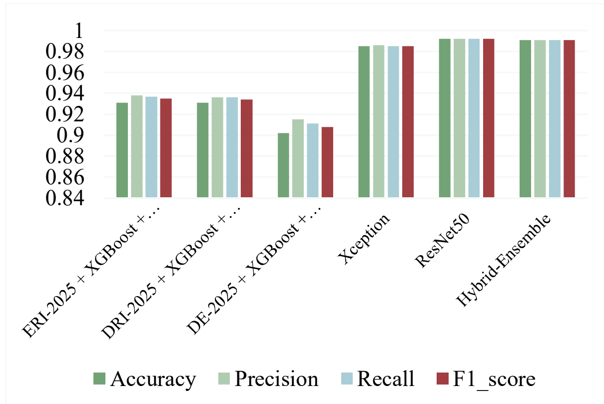
$$F1\ Score = 2 * \frac{Recall \text{ X } Precision}{Recall + Precision} * 100 \quad (1)$$

Table. 1. Performance Evaluation Table

ML Model	Accuracy	Precision	Recall	F1_score
ERI-2025 + XGBoost + SVM	0.931	0.938	0.937	0.935
DRI-2025 + XGBoost + SVM	0.931	0.936	0.936	0.934
DE-2025 + XGBoost + SVM	0.902	0.915	0.911	0.908
Xception	0.985	0.986	0.985	0.985
ResNet50	0.992	0.992	0.992	0.992
Hybrid-Ensemble	0.991	0.991	0.991	0.991

The performance evaluation table compares the different models with each other, showing that ResNet50 and Hybrid-Ensemble achieve the highest accuracy with balanced precision, recall and F1-score, exceeding the standard XGBoost-SVM combinations.

Fig. 2. Comparison Graph



The comparison graph shows the performance measures between models, with a higher accuracy, precision, recall, and F1-score of ResNet50 and Hybrid-Ensemble, but the variations of XGBoost-SVM show relatively lower and consistent results.

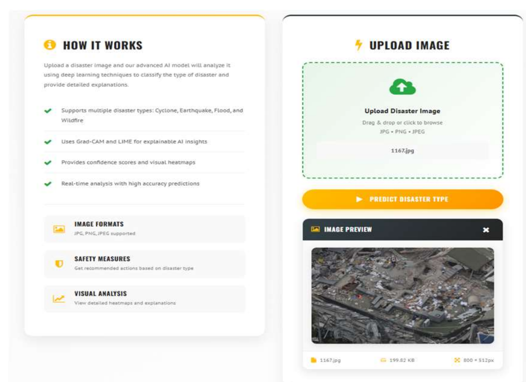


Fig.3 Upload the Image

Figure 2 illustrates the user interface of the catastrophe image input, which shows the upload process, preview of the image, and predict service used to categorize the type of disaster through the proposed AI-based system.

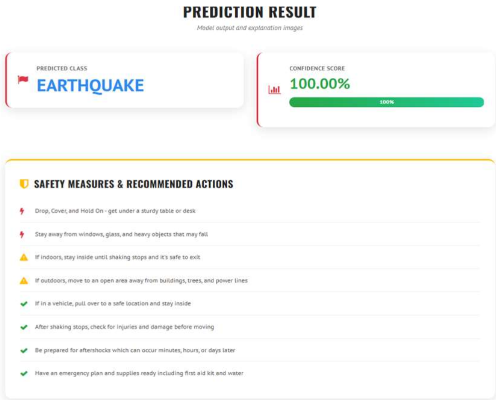


Fig.4 Predicted Result

Figure 4 shows the forecast result which points to an earthquake with an absolute confidence of 100.00 percent plus the safety measures and recommended actions to help the user take appropriate measures in case of such disasters.

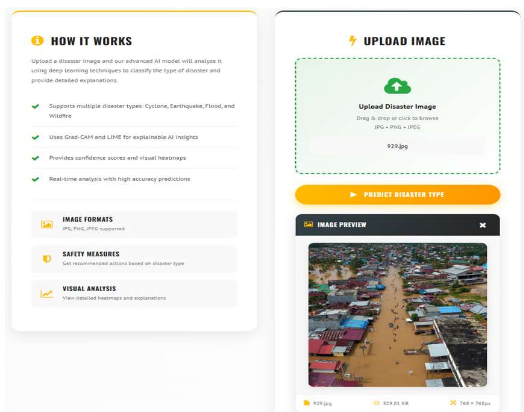


Fig. 5 Upload the Image

An aerial view of serious urban flooding in Figure 5 depicts a hazy water flooding streets and houses and is featured in the disaster prediction interface in the preliminary review before categorizing the disaster type.

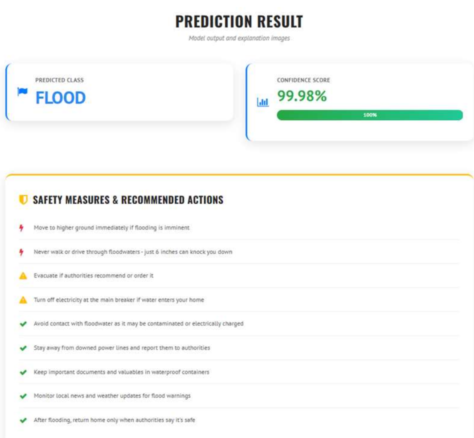


Fig. 6 Predicted Result

Figure 6 shows the forecast result that the disaster was a Flood with a confidence score of 99.98, along with recommended safety measures and steps of the disaster response and preparedness process to be effective.

V. Conclusion

The effective and timely classification of natural disasters in cities and rural areas is critical to strengthen the community and mitigate the impact of floods, cyclones, earthquakes, and wildfires. GeoDisasterAINet addresses this challenge by applying a multi-stage ensemble design to the Kaggle Natural Disaster Image Dataset which is composed of tagged images of various disaster events. The methodology uses the base models ERI-2025, DRI-2025, and DE-2025, which are supplemented with XGBoost, along with SMOTE and regular scaling to train fairly. The final step involves applying multiclass SVM to the XGBoost-enhanced models after standard scaling, and deep feature extraction using convolutional neural networks Xception, ResNet50, and a Hybrid Ensemble of Xception and ResNet50. The criteria used to evaluate the efficacy of the system include accuracy, precision, recall and F1-score and the results of the ResNet50 show that the test accuracy of the system is 99.2%. Enhancesments such as Grad-CAM and LIME provide explainable AI visualizations which highlight important areas of the image to influence predictions, and a Flask-based interface facilitates real-time deployment and user interaction. GeoDisasterAINet offers accurate, interpretable, and scalable disaster classification, which can support reliable decision-making and automated operational support of disaster management in cities and the countryside.

To enhance the system, additional disaster types, such as landslides, tsunamis, and droughts, may be introduced to enhance the geographic coverage in different geographic areas. Combination of multimodal sources of data like satellite images, sensor networks and real-time weather data would enhance the accuracy of the forecasts and situational awareness. The development of deep learning models, such as attention-based networks and transformer models, has the potential to increase the effectiveness of feature extraction and classification. Furthermore, deployment to cloud or edge computing devices can support large scale and real-time disaster monitoring. Better explainability methods and interactive dashboards can provide more intuitive data to decision-makers and allow acting swiftly and allocating resources in crisis management circumstances.

References

[1] Wen, L., Xiao, Z., Xu, X., & Liu, B. (2025). Disaster Recognition and Classification Based on Improved ResNet-50 Neural Network. Applied Sciences, 15(9), 5143.

[2] Chen, J., Seng, K. P., Lim, C. S., Smith, J., Ang, L. M., & Xu, H. (2025, June). A Distributed AI Framework for Enhancing Situation Awareness in Natural Disasters. In 2025 32nd International Conference on Systems, Signals and Image Processing (IWSSIP) (pp. 1-5). IEEE.

[3] Sufi, F., & Alsulami, M. (2025). AI-Driven Global Disaster Intelligence from News Media. Mathematics, 13(7), 1083.

[4] Kulahara, M., Kashyap, G. S., Joshi, N., & Soni, A. (2025). Can We Predict the Unpredictable? Leveraging DisasterNet-LLM for Multimodal Disaster Classification. arXiv preprint arXiv:2506.23462.

[5] Chakrabortty, R., Ali, T., Abouleish, M., Atabay, S., Ahmad, N., Abu-Rukba, R. A., ... & Al-Etoom, S. M. (2025). Urban Flood Susceptibility Assessment in Arid Environment Using a Novel Hybrid Deep Learning Approach. Earth Systems and Environment, 1-25.

[6] United Nations Office for Disaster Risk Reduction, The Human Cost of Disasters: An Overview of the Last 20 Years (2000–2019), Geneva,

Switzerland, 2020. [Online]. Available: https://www.preventionweb.net/publication/human-cost-disasters-overview-last-20-years-2000-2019

[7] Our World in Data, “Natural disasters,” 2023. [Online]. Available: https://ourworldindata.org/natural-disasters

[8] Centre for Research on the Epidemiology of Disasters (CRED), EM-DAT: The International Disaster Database, Brussels, Belgium, 2023.

[9] National Disaster Management Authority (NDMA), Annual Report on Disasters in India, New Delhi, India, 2023. [Online]. Available: https://ndma.gov.in/en/disaster-data.html

[10] S. K. Abid, N. Sulaiman, S. W. Chan, U. Nazir, M. Abid, H. Han, A. Ariza-Montes, and A. Vega-Muñoz, “Toward an integrated disaster management approach: How artificial intelligence can boost disaster management,” Sustainability, vol. 13, no. 22, p. 12560, Nov. 2021.

[11] W. Sun, P. Bocchini, and B. D. Davison, “Applications of artificial intelligence for disaster management,” Natural Hazards, vol. 103, no. 3, pp. 2631–2689, Sep. 2020.

[12] A. Rathod, V. Pariawala, M. Surana, and K. Saxena, “Leveraging CNNs and ensemble learning for automated disaster image classification,” arXiv:2311.13531, 2023.

[13] M. Rahman, M. S. Islam, and M. R. Islam, “Flood susceptibility mapping in Bangladesh using machine learning and multi-criteria decision analysis,” Geoenvironmental Disasters, vol. 11, Art. no. 150, Aug. 2024.

[14] C. Yu, B. Hu, and Z. Wang, “Open-world disaster information identification from multimodal social media,” Complex & Intelligent Systems, vol. 11, no. 7, pp. 1–16, Nov. 2024.

[15] M. A. Islam, F. Rabbi, and N. U. I. Hossain, “Performance evaluation of NLP and CNN models for disaster detection using social media data,” Social Network Analysis and Mining, vol. 14, no. 1, pp. 1–7, Nov. 2024.

[16] I. Alisjahbana, J. Li, and Y. Zhang, “DeepDamageNet: A two-step deep-learning model for multi-disaster building damage segmentation and classification using satellite imagery,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2024, pp. 1–14.

[17] R. Umeike, T. Dao, and S. Crawford, “Accelerating post-tornado disaster assessment using advanced deep learning models,” in Proc. IEEE MetroCon, Nov. 2024, pp. 1–3.

[18] B. Jankovic, S. Jangirova, W. Ullah, L. U. Khan, and M. Guizani, “UAV-assisted real-time disaster detection using optimized transformer model,” arXiv:2501.12087, 2025.

[19] S. K. Bhowmick, A. Barman, and S. K. Roy, “Thresholding metrics for evaluating explainable AI models in disaster response,” Digital Signal Processing, vol. 160, Art. no. 105068, May 2025.

[20] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic minority over-sampling technique,” J. Artif. Intell. Res., vol. 16, pp. 321–357, Jun. 2002.

[21] M. T. Ribeiro, S. Singh, and C. Guestrin, “‘Why should I trust you?’ Explaining the predictions of any classifier,” in Proc. NAACL-HLT, San Francisco, CA, USA, 2016, pp. 1135–1144.

[22] Kaggle, “Natural Disaster Image Dataset,” Jan. 2025. [Online]. Available: https://www.kaggle.com/datasets/alex1994/natural-disaster-image-dataset

[23] H. Wu, Q. Huang, D. Wang, and L. Gao, “A CNN-SVM combined model for pattern recognition,” J. Electromyography Kinesiology, vol. 42, pp. 136–142, Oct. 2018.

[24] T. Zhao, X. Zhang, and S. Wang, “Imbalanced node classification with synthetic over-sampling,” IEEE Trans. Knowl. Data Eng., vol. 36, no. 12, pp. 8515–8528, Dec. 2024.

[25] J. Wang and N. Awang, “A novel synthetic minority oversampling technique for multiclass imbalance problems,” IEEE Access, vol. 13, pp. 6054–6066, 2025.

[26] M. Tan and Q. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in Proc. Int. Conf. Mach. Learn., Sep. 2020, pp. 6105–6114.

[27] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, “Inception-v4, inception-resnet and the impact of residual connections on learning,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 1–7.

[28] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jul. 2017, pp. 2261–2269.

[29] S. Biswas, R. Mostafiz, M. S. Uddin, and B. K. Paul, “XAI-FusionNet: Diabetic foot ulcer detection based on multi-scale feature fusion with explainable artificial intelligence,” Heliyon, vol. 10, no. 10, Art. no. e31228, May 2024.

[30] O. Rainio, J. Teuho, and R. Klén, “Evaluation metrics and statistical tests for machine learning,” Scientific Reports, vol. 14, no. 1, p. 6086, Mar. 2024.