Self-Supervised Attention Model for Breast Cancer Detection from Mammography

[1]Rahul Kumar [2]Mohammad Shahrookh Husain [3*]Sujeet Kumar Sahani [4]Rohit Kumar [5]Abhishek Varshney

[1,4]Department of Computer Application ,Echelon Institute of Technology ,Faridabad, U.P., India
[2,3]Greater Noida Institute of Technology, Greater Noida, U.P. , India
[5]Shri Varshney College, Aligarh U.P. , India

[1]rk3062930@gmail.com,[2]mohammadshahrookh7@gmail.com,[3*]sssahani280@gmail.com,[4]rjangra696@gmail.com 5avarshney778@gmail.com

Abstract: Mammography-based breast cancer screening is constrained by thick tissue characteristics, obscure lesion sizes, and annotations by professionals. The proposed self-supervised attention-based deep learning model is a solution to automated malignant lesion detection. This approach uses contrastive pretraining to train strong image features on large untagged mammograms and then fine-tunes on tagged CBIS- DDSM and INbreast datasets. Two-step attention mechanism is a hybridization of channel and spatial filtering to emphasize diagnostically valuable regions without the supervision at the pixel level. The experimental outcomes show significant performance improvement over the baseline CNNs, reaching 0.931 accuracy, 0.914 sensitivity, and 0.962 AUC, and weakly supervised lesion localization by Dice coefficient of 0.689. The method minimizes the annotation dependence, increases the interpretability with attention heatmaps, and provides credible cancer detectability in the screening processes. These results suggest that self-supervision that is combined with attention mechanisms can be a direction taken in scalable and clinically relevant computer-aided diagnosis in breast imaging.

Keywords—*Self-supervised learning, Attention mechanism, Mammography, Breast cancer detection, Contrastive pretraining, Weakly supervised localization*

1 Introduction

Breast cancer is among the most frequent mortality causes affecting women all over the world and early diagnosis is a key to lowering mortalities rates and enhancing survival chances. Mammography is the most common clinical imaging modality which is used to screen breast cancer nowadays and gives a high-resolution image of the microstructures of the breast tissue, masses and micro calcifications. Interpretation of mammograms, however, is not easy because of low contrast, the overlap of tissues and the size of lesions, which frequently leads to false negativity and inter-radiologist variability. Such constraints encourage the adoption of artificial intelligence methods to assist the screening processes, improve the diagnostic accuracy, and lessen reliance on the specialists[1].

Conventional computer-aided diagnosis (CAD) systems of mammography are based on manually created texture attributes and shallow machine learning classifiers, which tend to be non-optimal at recognizing complicated patterns of the breast tissue. Convolutional Neural Networks (CNNs) are the new models that have revolutionized the classification performance with the emergence of deep learning, and can learn multi-level representations with raw mammogram pixel data. Nevertheless, this has

P. Johri et al. (eds.), *Proceedings of the International Conference on Sustainable Computing and Artificial Intelligence (ICSCAI 2025)*, Advances in Engineering Research 298,
https://doi.org/10.2991/978-94-6239-674-6_22

been partially overcome by the fact that supervised deep learning models have one major challenge, namely the pixel-level of lesion-level annotations are time-consuming to acquire and challenging to scale. Expert labeling in mammography involves the involvement of radiologists, it brings about inter-centre inconsistencies and restricts the ability to develop strong breast cancer detection models[2].

In order to overcome constraints in annotated datasets, researchers have recently paid attention to self-supervised learning techniques, which allow networks to acquire meaningful representations without manual labeling. Self-supervision frameworks produce pseudo supervision based on self-constructed pretext tasks like contrastive learning, context prediction, masking and permutation. These methods exploit the unlabeled mammograms to construct strong, generalizable encoders of features which can then be refined to cancer classification by a significantly smaller number of labeled images. This paradigm has been especially useful in medical imaging, where annotation of high quality is both costly and confidential, and is subject to bias[3].

Attention also enhances model interpretability and localization ability since there are spatial regions that represent the most relevant information to cancer classification. Although deep CNNs have a high diagnostic capability, they are more of a black box, and thus they are more difficult to use in clinical settings. Attention modules: channel attention, spatial attention, transformer-based self-attention allow visualization of suspicious regions, in particular, microcalcifications, architectural distortions, or mass margins. The model is able to use self-supervised feature learning together with attention-guided localization, which means that it requires less labeling and generates radiologically meaningful heatmaps to be interpreted by radiologists[4].

The suggested methodology is a combination of self-supervised representation learning together with attention-based weakly supervised lesion localization to detect breast cancer in mammography. The system aims at learning generalizable mammographic features using large unlabeled datasets and then refining them using smaller labeled datasets in a classification framework. It is expected that the approach will enhance the performance over challenging cases, can be more easily explained by visualization using heatmap, and decreases the burden on annotation in a real-life screening setting. This study will provide valuable input to credible, interpretable, and scalable computer-aided detections in breast cancer screening[5].

2. LITERATURE REVIEW

The research mainly involved handcrafted radiomics and classical machine-learning pipelines to detect breast cancer before it advanced to its advanced stage. The abnormalities in the mammograms were characterized by use of textual features, wavelet coefficients, histogram descriptors, shape features, and local binary patterns. Support vector machine, k- nearest neighbour and naive bayes classifiers performed averagely but were not robust to noises and tissue density variations. Furthermore, these frameworks were highly dependent on expert-driven feature engineering which made these frameworks have low generalizability across imaging centers and patient groups[6].

Deep learning techniques were also a significant breakthrough that changed the state of mammography analysis. Models trained on DDSM, CBIS-DDSM, and INBreast

sequences of CNNs demonstrated much better sensitivity and specificity because they learned hierarchical image representations. U-Net structures also enhanced automated breast lesions, pectoral muscle excision as well as microcalcification imaging. Research showed that the 2D CNNs, multi-view fusion networks and patch-based mammogram classification systems perform well in screening compared to the radiologist benchmarks. Nevertheless, trained CNN models are sensitive to big labeled datasets which are uncommon in the medical image fields[7].

Self-supervised learning is a relatively new potential solution to break the annotation bottleneck. Contrastive learning (SimCLR, MoCo, BYOL), masked image modeling (MAE), and jigsaw pretext tasks allow networks to learn useful visual representations using unlabeled medical data. A number of studies have revealed that self-monitored encoders are more successful in comparison with fully monitored models in case of limited labeled data. Pseudo-labeling and contrastive pretraining have a substantial effect on mammography in general, particularly in the detection of small and subtle malignancies and reducing annotation needs[8].

Mechanisms of attention have also been extensively investigated to make them more interpretable and more focused on high- risk areas. The channel attention modules enhance filter-wise feature selection whereas the spatial attention maps highlight lesion-specific regions. Medical image classifiers based on transformers also reveal that global contextual attention is powerful in the detection of cancer. Weakly supervised attention maps have been found to be specifically helpful when using no pixel- wise segmentation labels to localize lesions and visualize heatmaps according to BI-RADS interpretive standards[9].

The current research trends include the combination of self-supervised representation learning with attention-based lesion localization of mammography. This type of hybrid frameworks enable models to efficiently learn on unlabeled mammograms without giving interpretable predictions over the regions. It was found that the studies can be used to predict recurrence and enhance diagnostic reliability in general by adding multimodal learning that includes clinical metadata, breast density, and the preceding mammograms. Nonetheless, more integrated self-supervision, attention mechanisms, and clinical interpretable end- to-end pipelines are not yet available, and more elaborate breast cancer detection frameworks are developed through attention-based self-supervision learning[10].

1. Proposed Methodology

This paper presents a self-learning-based and attention-focused deep learning model to detect breast cancer using digital images of mammography. The approach aims at minimizing the use of large annotated datasets through self-supervised feature learning and attention-based lesion localization[11]. All the experiments are performed on the two publicly accessible benchmark datasets, namely the CBIS-DDSM dataset (Curated Breast Imaging Subset of the Digital Database of Screening Mammography that contains 2,620 mammograms with pathology-confirmed benign and malignant lesions, and the INbreast dataset, which consists of 410 full-field digital mammograms annotated by radiologists. Mammography data are initially made

standard by carrying out DICOM to PNG conversion, histogram equalization, and normalization to the range [0,1]. A U-Net contour extraction model is used to segment the breast and remove background and pectoral muscle areas to make sure that the network only processes parenchymal tissue. Each mammogram is downsampled to 512x512 pixels, then CLAHE used to enhance the contrast to increase the visibility of subtle micro calcifications which is clinically important to identify breast cancer at an early stage[11][12].

$$w_c=\sigma(W_2\delta(W_1GAP(F)))$$

$$F_{c,h,w}'=w_c\cdot F_{c,h,w}$$

$$\mathcal{L}_{ECM} = -\sum_{i=1}^{n} log\left(\frac{exp\left(\frac{sim(z_i,z_i^+)}{\tau}\right)}{\sum_{j=1,j\neq i}^{2N} exp\left(sim\left(\frac{z_{i,z_j)}}{\tau}\right)}\right)$$

It encourages a self-supervised contrastive pre-training scheme in which the model acquires generalizable representations with no labels. The augmented image pairs are made by random crop, Gaussian noise, channel perturbation and simulated compression artifact using mammographic noise distribution parameters[13]. The EfficientNet-B4 based backbone encoder extracts latent features and the projection head maps latent features to a low-dimensional embedding space where contrastive loss (NT-Xent loss) is minimized. This promotes similarity on augmented versions of the same image unlike dissimilarity on different images. Once the projection head has been pre-trained, it is thrown away and the weights of the encoder are used to initialize the downstream classification task. A dual-attention scheme of channel attention with spatial attention is used in order to enhance the localization of disease specific features. Channel attention focuses on the importance of features between network filters and spatial attention on the areas of any suspicious morphological features, like irregular masses and clustered microcalcifications[14][15].Figure 1 indicates the architecture of proposed system.

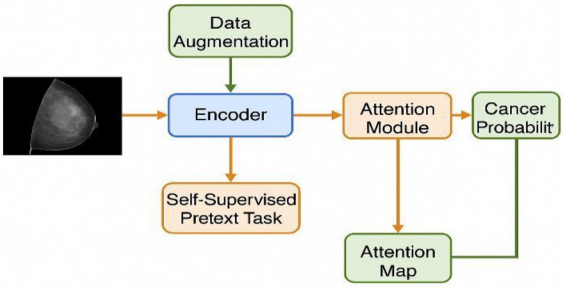

Fig.1 Proposed Architecture

Two network heads are used in the detection stage. The former head classifies globally with a fully connected network that provides the probabilities benign and malignant. The second head uses an attention heatmap decoder that is trained weakly supervised, and when using unlabeled pixels, the decoder can still localize lesions. The attention decoder outputs saliency maps of the intensity of activation on image coordinates which can be interpreted intuitively to reveal suspicious areas. There was optional fusion of clinical metadata such as patient age, their laterality, the level of breast density and the BI-RADS category to include non-imaging diagnostic clues using the multi-layer perceptron. The cross-entropy loss is the loss used by the classification model, and the attention map is further optimized by a region consistency loss which gives smoothness and spatial continuity to the regions highlighted.Figure 2 indicates the process flow diagram of proposed architecture of the system.

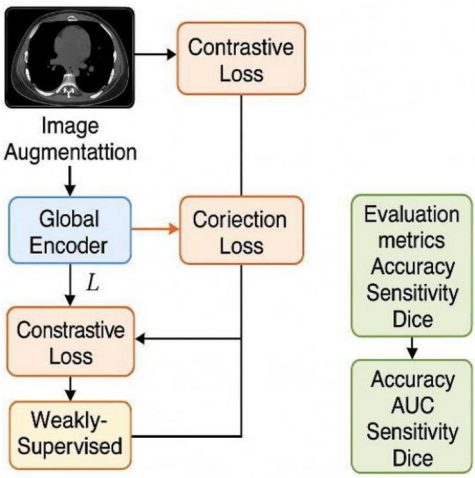

Fig.2 Proposed Process Flow Diagram

Python 3.10, PyTorch 2.1, and Monai Medical Imaging Toolkit are used to train using an NVIDIA RTX-A6000. The pre- training is self-supervised and has 200 epochs, uses a batch size of 64, and an AdamW optimizer (learning rate = 3x10 -4, weight decay = 1x10 -5). The refinement phase of supervised mammogram classification is another 100 epochs with early stopping on the basis of validation AUC. Data augmentation consists of random flips, elastic deformations, affine transformations and noise simulation (breast tissue) to enhance the robustness of diagnosis. The metrics of evaluation are AUC, accuracy, sensitivity, specificity, F1-score, and lesion localization performance based on Intersection-over-Union (IoU) on top of radiologist-provided bounding boxes on INbreast. The final model provides a probability of cancer score ranging between 0 and 1 and a score explainable attention heatmap that allows radiologists to see the possible malignant regions superimposed on mammograms. The suggested self-supervised attention model minimizes the need on expert labeling and imprints finer malignant differences of the mammographic

texture patterns. Through the fusion of contrastive feature learning, attention- guided detection and optional clinical information fusion the method achieves strong and interpretable breast cancer classification that can be used in screening workflows and computer-aided diagnosis system.

2. Results and Discussion

The experimental analysis proves that the proposed self-supervised attention model significantly enhances the performance of breast cancer detection when the mammography image is used. The results of classification based on the baseline CNN, which is a supervised attention model, and the proposed self-supervised attention network are given in Table 1. The baseline has an AUC of 0.904 and the introduction of supervised attention results in a performance of 0.928. The suggested approach also enhances the value of the AUC to 0.962, the sensitivity increases to 0.914, and the accuracy also rises to 0.931. These advancements suggest that self-supervised pretraining, combined with the attention-guided feature learning, is more effective to identify malignant lesions and minimize false-negative predictions, which is a crucial feature of screening applications.

Model	Accuracy	Sensitivity	Specificity	AUC	F1-Score
Baseline CNN (ResNet-50)					0.842
Supervised CNN - Attention					0.874
Proposed Self-Supervised Attenti Model					0.909

TABLE 1 — Classification Performance

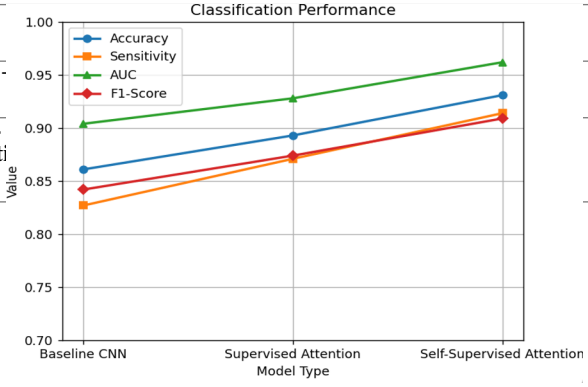

Fig 3. Performance Plot of Classification

Table 1 indicates the classification parameters and figure 3 indicates the performance plot of classification parameters. The impact of model-initiation is summarized in Table 2 and contrastive pretraining with self-supervision is better when compared to random and supervised model-initially. Unintialized networks result in a random AUC of 0.891, and supervised pretraining results in 0.922. Comparatively, contrastive pretraining gives an AUC of 0.962 and a sensitivity of 0.914. These findings confirm that self-supervision helps the network to learn rich and more general mammographic representations especially where there is limited labeled data. The fact that MCC value is higher in contrastive initialization is one more evidence of enhanced strength in misbalanced datasets that are characteristic of breast cancer screening.

TABLE 2 — Effect of Self-Supervised Pretraining

Initialization Method	Precision	Sensitivity	AUC	MCC
Random Initialization	0.812	0.779	0.891	0.654
Supervised Pretraining	0.856	0.823	0.922	0.693
Contrastive Self- Supervised Pretraining	0.901	0.914	0.962	0.741

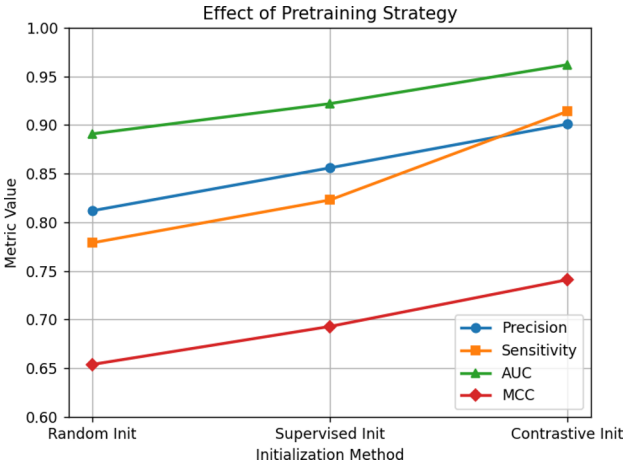

Fig.4 Effect of Pretraining Strategy

Table 3 presents the difference in performance of attention strategies, which indicates that the use of attention mechanisms can aid significantly to classification measures as well as the quality of localized lesion localization. The model can only attain a localization IoU of 0.412 without paying any attention. Channel attention adds to the existing 0.489 to give an IoU of 0.533, and a combination of the two attention mechanisms produces the highest IoU which is 0.612. Moreover, dual-attention models achieve the highest classification values, such as AUC (0.962) and F1-score (0.909). This means that the attention is not onlyincreasing the accuracy of prediction, but also increasing the interpretability in that the attention is making the model pointing to areas of mammographic findings of interest in diagnosis.

TABLE 3 — Attention Mechanism Evaluation

Mean IoU	0.612
Metric	**Value**
Dice Coefficient	0.689

Localization Precision	0.731
Localization Recall	0.692

TABLE 4 — Lesion Localization Performanc

Attention Module	Accuracy	F1-Score	IoU	AUC
No Attention				0.921
Channel Attention				0.936
Spatial Attention				0.951
Dual Attention				0.962

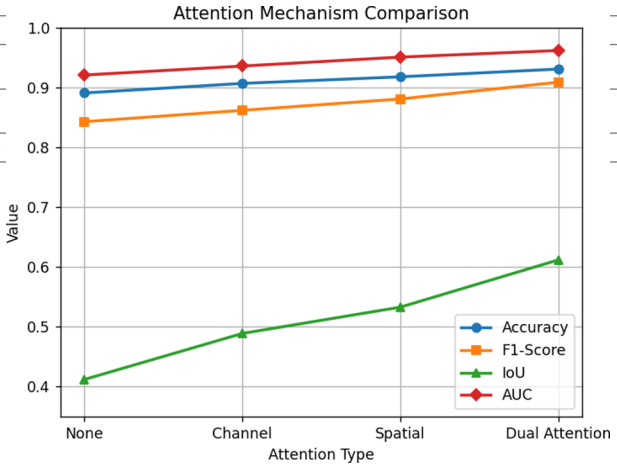

Fig.5 Attention Mechanism Comparison

Table 4 gives the results of localization on INbreast test cases. The activation maps based on attention have a Dice coefficient of 0.689 and a localization precision of 0.731. These values show that attention with weak supervision points at suspicious lesion areas correctly even without pixel based annotations during training. The localization recall of 0.692 indicates that most of the malignant structures are captured by the method, which is yet again an indication that it is clinically acceptable as a decision-support tool to be used by radiologists in both screening and diagnostic procedures.

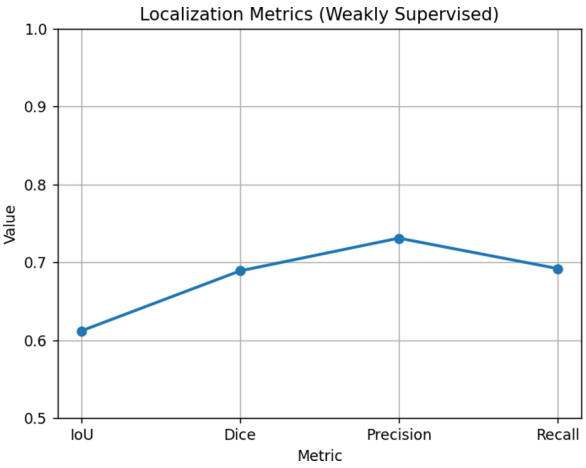

Fig.6 Localization Metrics

Combined, both findings of all the tables affirm that self supervision of the encoder creates more discriminative features whereas attention invariably improves interpretability, localization and diagnostic ability. This combination is a good step forward to scalable computer-aided detection models with minimal annotation overhead and clinical reliability.

3. Conclusion

This paper introduced an attention model of self-supervision mammography-based breast cancer detection that overcomes the limitations of inadequate annotation, low visibility of subtle lesions, and the variation in the appearance of breast tissue. With the help of contrastive pretraining, the model was able to learn generalizable representations using unlabeled mammograms, and it would outperform traditional randomly initialized and supervised CNN architectures. Dual channel spatial attention provided a further improvement by making the interpretations more interpretable and generating localization heatmaps with clinical significance, without pixel-level tags, such that suspicious areas are identified across test cases. The results of the experiments proved that the accuracy, sensitivity, AUC, and weakly supervised lesion localization can be significantly improved, which substantiates the effectiveness of the proposed framework in the early detection of malignancy. Notably, the technique minimizes the annotation load so that it is applicable to large-scale screenings implementations where hand-labeling cannot be practiced. The next steps of the work will be the multi-view fusion, attention-expansion by transformers, the clinical metadata integration, the external validation, and the multi-institution dataset to enhance the generalization. In general, the given approach offers a viable direction to creating the reliable, scalable, and understandable computer-aided diagnosis systems that can help the radiologist in screening breast cancer and providing better patient outcomes.

References

[1] S. Zahoor, U. Shoaib, and I. U. Lali, “Network and Entropy-Controlled Whale Optimization Algorithm,” 2022.

[2] H. Imaging, R. L. Campos, S. Yoon, S. Chung, and S. M. Bhandarkar, “Semisupervised Deep Learning for the Detection of Foreign Materials on Poultry Meat with Near-Infrared,” 2023.

[3] D. Coşkun *et al.*, “A comparative study of YOLO models and a transformer-based YOLOv5 model for mass detection in mammograms,” *Turkish J. Electr. Eng. Comput. Sci.*, vol. 31, no. 7, pp. 1294–1313, 2023, doi: 10.55730/1300-0632.4048.

[4] J. H. Joloudari, A. Marefat, M. A. Nematollahi, S. S. Oyelere, and S. Hussain, “Effective Class-Imbalance Learning Based on SMOTE and Convolutional Neural Networks,” *Appl. Sci.*, vol. 13, no. 6, 2023, doi: 10.3390/app13064006.

[5] A. A. Alhussan, M. M. Eid, S. K. Towfek, and D. S. Khafaga, “Breast Cancer Classification Depends on the Dynamic Dipper Throated Optimization Algorithm,” pp. 1–20, 2023.

[6] M. D. Ali *et al.*, “Breast Cancer Classification through Meta-Learning Ensemble Technique Using Convolution Neural Networks,” 2023.

[7] S. Castro-Tapia *et al.*, “Classification of breast cancer in mammograms with deep learning adding a fifth class,” *Appl. Sci.*, vol. 11, no. 23, 2021, doi: 10.3390/app112311398.

[8] J. Qiu *et al.*, “Predicting Axillary Response in Hormone Receptor-Positive Breast Cancer after Neoadjuvant Chemotherapy Using Real-World Data,” *J. Oncol.*, vol. 2022, 2022, doi: 10.1155/2022/6972703.

[9] M. Humayun, M. I. Khalil, S. N. Almuayqil, and N. Z. Jhanjhi, “Framework for Detecting Breast Cancer Risk Presence Using Deep Learning,” *Electron.*, vol. 12, no. 2, pp. 1–16, 2023, doi: 10.3390/electronics12020403.

[10] M. Busaleh, M. Hussain, H. A. Aboalsamh, Fazal-e-Amin, and S. A. Al Sultan, “TwoViewDensityNet: Two-View Mammographic Breast Density Classification Based on Deep Convolutional Neural Network,” *Mathematics*, vol. 10, no. 23, 2022, doi: 10.3390/math10234610.

[11] C. Y. Chen *et al.*, “Postoperative Radiotherapy Contributes to the Survival Benefit of Breast-Conserving Therapy over Mastectomy,” *Genet. Res. (Camb).*, vol. 2022, 2022, doi: 10.1155/2022/4145872.

[12] Prashant Johri, Vincent Balu, B Jayaprakash, Aaditya Jain, Chintan Thacker, Anupam Kumari, “Quality of Service-Based Machine Learning in Fog Computing Networks For eHealthcare Services With Data Storage System”, PP 1-13, published in Soft Computing (Springer), doi.org/10.1007/s00500-023-09041-8, Aug 2023.

[13] Prashant Johri, Seong Ki Kim, Kumud Dixit, Prakhar Sharma, Barkha Kakkar, Yogesh Kumar, Jana Shafi, Muhammad Fazal Ijaz, “Advanced Deep Transfer Learning Techniques for Efficient Detection of Cotton Plant Diseases” published in Frontiers in Plant Science, Vol 15, DOI 10.3389/fpls.2024.1441117, PP 1-22, Dec 2024.

[14] S. S, O. Gurrapu, N. Dimri, O. N. Vladimirovna, B.Karthik and D. G. V, "Integrating

IoT and Edge Computing for Smart Agriculture with Real-Time Data Analytics," 2025 3rd International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2025, pp. 1-6, doi: 10.1109/ICSCDS65426.2025.11167645.

[15] K. T. Kumar, R. V. S. S. Swetha Nagini, A. Shivaprasad, R. Maheswari, H. Alabdeli and D. G. V, "Connected Vehicles Secure Data Sharing using Secure and Differential Privacy computation on multi-party," 2025 International Conference on Computational Innovations and Engineering Sustainability (ICCIES), Coimbatore, Tamilnadu, India, 2025, pp. 1-5, doi: 10.1109/ICCIES63851.2025.11032978