[1] R. Prasanna Kumar

[2] Rohini V

[3] Kavitha Subramani

[4] G. Malathi

[5] P. Jagadeesan

Development of Multi Modal Image Fusion Techniques in Night Vision Technology Using Deep Learning

JES

Journal of Electrical Systems

Abstract: - This study is aimed at exploring new development in nighttime image processing based on nighttime photography/night vision. We use various learning methodologies based on neural networks to develop multisensor data processing techniques. The ability of Convolutional Neural Networks (CNN) to enhance image quality and clarity under conditions of low light or at night is the focus of our research. To do this A large database has been created, containing 3200 images both taken from the internet as well as our own. This data is preprocessed, normalized and features expanded using some common methods of data augmentation in order to optimize model performance. We compared our proposed CNN model with existing literature through a series of experiments. Performance was measured using Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Feature Similarity Index (FSIM). The Results are now in. They showed that our model can improve various metrics with higher scores. And they show the real-time monitoring capabilities in urban streets for our model. This research brings multisensor image fusion based on night-adjacent techniques forward from night vision technology and provides insight into practical and potential future directions for using deep learning techniques to improve safety and security in many different environments.

Keywords: Multi-modal image fusion, Night vision technology, Deep learning, Surveillance, Security.

I. INTRODUCTION

Multi-modal image fusion is the key of night vision technology, the mixing of different sensors, to upgrade image quality, clarity, and comprehensibility in low-light or nighttime conditions. Here we will give a comprehensive review of published papers about multi-modal image fusion method focusing on both traditional techniques from 60 years ago till now, and some recent innovations due to deep learning. Because night vision technology tends to have many applications both military and civilian Weighted averaging is the most elementary method: calculate a weighted average pixel intensity together from sensor modalities. Although simple weighted averaging is easy to carry out, it will not effectively capture complex relationships between image modalities, which tends to leave room for much room for improvement [1]–[3]. Principal Component Analysis (PCA) is another widely used technique, employing dimensionality reduction to combine pictures by retaining the most representative orthogonal components. As PCA-based fusion methods may effectively capture the most relevant information from multimodal images, there can be information loss and interpretation loss. Wavelet Transform decomposes images into different frequency components and uses such rules as maximum, minimum and average to fuse them together. Wavelet fusion approaches are somewhere in between – computational efficiency and visual outcome are almost equally good [4]–[6].

[1] *Corresponding author: Department of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Chennai, India. Email: kumarprasanna.r@gmail.com, ORCID: 0000-0001-6103-259X

[2] Associate Professor, Department of MBA, SA Engineering College, Chennai. Email: rohiniv@saec.ac.in, ORCID: 0009-0008-3973-5238

[3] Professor, Department of CSE, Panimalar Engineering College, Chennai. Email: kavitha.pec2022@gmail.com,

ORCID: 0000-0003-4565-1029

[4] Professor, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai. Email: malathi.g@vit.ac.in,

ORCID: 0000-0002-0303-7402

[5] Assistant Professor, Department of Computer Science and Engineering, R.M.D Engineering College, Kavaraipettai 601206.

Email: pjn.cse@rmd.ac.in

Multi-modal image fusion processes have been redesigned from ground up by deep learning, and with highly developed tools experts barely have to do any of the hard work now. Convolutional Neural Networks (CNNs) were the dominant architectural design for multi-dimensional image fusion owing precisely to the way they could take figures and run them through many layers processing, each interpreting some aspect in might eventually become this composite picture you see when everything processed together at once. CNN-based fusion methods often train the neural network on matched multi-sensor data, then use loss functions tailored to a specific objective, like increasing similarity or reduction in reconstruction error [7], [8]. Generative Adversarial Networks (GANs) represent a new way of tackling this issue." If a generator network can learn how to make the joint product of two inputs with different modalities -- then generate realistic pictures, discrinator networks are set up to evaluate their output. GAN-based methods can lead to really beautiful results unless you need visual appeal; but they have higher energy requirements and the hyperparameters need more thorough consideration [9], [10].

Featuring Siamese Networks and Autoencoders alongside CNNs and GANs, we investigate a total of six algorithms handled under deep learning architectures for multi-modal image fusion. Siamese Networks are a learning machine that measures the similarity essays between images, and for multi-modal fusion applications can also generate joint pairings of their different modalities. Autoencoders are used in unsupervised learning of features and for image reconstruction. Therefore, with deep-learning-based methods it is easier to find unsupervised features and create new composite images of high quality [11]–[13].

While multi-modal image fusion techniques have advanced significantly, there are several obstacles and constraints that must be met. One of the big challenges is that deep learning models require large-scale annotated datasets to train them well-performing. This is especially difficult in the area of night vision applications, where data with labels may be scarce and costly. In addition, deep learning-based fusion methods are not very interpretable. It is difficult to understand the influences performance or to recognize potential failure modes because of low interpretability. Moreover, in real-time applications the computational complexity of deep learning models may have practical limitations. Efficient model architectures and optimization techniques can also solve this problem [14]–[16].

The multi-modal image fusion techniques are essential in night vision technology because they improve our ability to see and interpret visual information under low light or dark conditions. Conventional fusion methods provide a firm footing for integrating information from different sensor modalities, while the latest methods of deep learning give us powerful ways to learn complex relationships among image modality types and produce high-quality fused images. Future lines of research in multi-modal image fusion might center on these challenges, including dataset availability, interpretability, and computational efficiency to eventually strengthen the capabilities of night vision in various practical applications [17], [18].

II. METHODOLOGY

In the field of surveillance, particularly in low-light or night-time conditions, it is necessary for security needs that multiple images taken from various kinds of sensors such as infrared (IR) and visible light cameras are integrated. However, the role of conventional surveillance systems is limited due to having low resolution. In the absence of good images under such conditions, one's sight is compromised and so is the effect of the system in identifying objects, individuals, activities, or movements within the monitored area. For security reasons, these limitations mean serious consequences as threats and suspicious activities go unnoticed possibly due to the poor quality of images.

The main driver for this research is the implementation of low-light surveillance systems to work around these defects and upgrade their usefulness. Combining images from various sensor modalities including visible camera and IR camera, the combined image is clearer and offers greater contrast and detail under severe are-conditions than ever before. The operator is given the ability to pick out critical information in every conceivable aspect, so judgments can be made on the spot only on self-evident. He also has greatly increased levels of situation awareness because of this improved image quality. This intensification of environmental awareness and efficiency in working with the environment greatly improves response.

The fusion of multi-modal images seems likely to have tremendous potential, but there are many significant difficulties that must be dealt with. These include developing robust fusion techniques that can successfully combine information from different sensor modes without losing any important characteristics. At the same time

they must also be able to suppress artifacts. Moreover, one will need efficient and 'real-time' fusion algorithms that can be included in a surveillance system whether used in a dynamic environment or not. Lastly, these methods must be adaptable and scalable for diverse surveillance scenarios and hardware platforms. This is yet another serious challenge.

On this foundation, the current study aims to create new multi-modal image fusion techniques that can significantly improve on today's poor performance with deep learning. The goal is to create innovative fusion methods that improve image quality and object detection and recognition by leveraging the power of deep learning algorithms. Incidentally, while in the dark and watching systems are ideal, this research also investigates systematic experiments and evaluations to produce not only publications but also practical designs. These have a wide range of potential applications, including internal security, defence, and civil law enforcement.

The main purpose of this study is to devise and appraise deep learning-based multimodal image fusion techniques that will enhance night vision technology. Specifically, it hopes to counter the drawbacks of taking detailed and clear images in low light or night-time conditions by consolidating information from different sensor modalities on a single platform, such as combining data obtained by both an infrared (IR) camera and visible light photography.

Several major goals are being sought by the research: First, to develop fusion models based on deep learning capable of uniting more imaginatively different types of information from irregular sensor data at the same time preserving the essential information features on an image. Second, to conduct a wide variety of experiments to evaluate the clarity, contrast changes, and detail preservation effect of this proposed fusion method.

Furthermore, it wants to explore in particular the operational efficiency and practicality of these developed fusion techniques in a range of real-time applications and under different surveillance settings. Moreover, it aims also to consider how the suggested techniques can help in the detection, recognition, and understanding of objects in such an environment.

The dataset of this research includes 3200 images from the Internet. The sources are visible light cameras and infrared sensor images. Using this dataset as a base, the deep-learning methodological judges can go ahead and train multi-modal picture fusion technology. Several preprocessing steps were taken in this research, leading to a collection of 3200 images from internet resources revealing light and infrared camera images.

III.   PREPROCESSING OF IMAGES

The first step in image data preprocessing would be data normalization. This entailed transforming pixel values across the whole dataset into a common interval, usually from zero-to-one. Normalization was conducted to reduce pixel intensity differences among images, making sure that the deep learning model gets is uniform input data. By normalizing the pixel values, the model could fully bring to bear during training phase and extract meaningful features across different sensor modalities. Furthermore, normalization helped the optimization process remain stable in that it kept too large (or small) gradients from plaguing all layers optimizing again to other local minima. This enhanced the overall robustness and convergence properties of the deep learning framework.

To make the deep learning models more intelligent, they would use data augmentation. This meant adding other examples to the dataset so that while learning, we could learn more things about that it could perceive in different contexts. Augmentation techniques included random rotations, translations, and changes in brightness and contrast for the training of images. These changes, based on randomness, were performed uniquely for each image during the training process; to thereby obtain many differences in appearance while preserving semantic content. Data augmentation had several purposes: exposing the model to more types of visual variations would help to avoid overfitting, and also saw a general improvement of performance, and supported the model in dealing with real-world variations such as illumination and perspective. Furthermore, data augmentation additionally had a regularizing effect on the model. This would result in more robust and general feature representations. The preprocessing stage consists of normalizing the data and conducting data augmentation together. Any training data was prepared as a way of ensuring that the dataset was already fitted for deep learning experiments on multimodal image fusion.

IV.   MACHINE LEARNING MODEL

In our research, we used Convolutional Neural Networks (CNNs) as the foundation for multi-modal image fusion in night vision technology. This type of CNN tipping point was a U-Net with a modified architecture, which combined well-known information from various sensor types. It consisted of an encoder-decoder network with skip connections to preserve spatial detail and multi-scale features during the fusion process. The first half of this neural network consisted of a number of convolutional layers, followed by max-pooling operations, which allowed it to compensate for space compression while increasing the number of feature maps. This collection of convolutional layers forms a hierarchy that can provide a network with hierarchical representations of both infrared and visible light images. The decoder portion of the network has the same structure as the encoder and is based on both upsampling operations to recover spatial dimensions bit by bit and generating fused output images. Following each convolutional layer, ReLU (Rectified Linear Unit) activation functions were used to add nonlinearity to the nonlinear transformation and complex feature representations that could be learned. The network contained batch normalisation layers that helped to stabilise and speed up the training process. This ensured that the network learned more accurate and common feature representations. The fused image was produced using a linear activation in the network's final output layer. During training, the loss function was a combination of mean squared error (MSE) loss and perceptual loss; differences in pixel values were penalised, but the network aimed to encourage high similarity between the fused image and the ground truth.

The Adam optimizer was used to optimise with the default parameters. It utilised its adaptive learning rate, which enables high-dimensional parameter space navigation to move quickly towards optimal solutions. A network of 3200 Internet images were trained end-to-end using an extensive dataset and an exponentially decaying learning rate schedule. The batch size was 32. After numerous experiments and careful hyperparameter tuning, the CNN architecture was shown to be quite adept at combining images from various sensors and styles to significantly improve image quality and allow for better detection and identification under the same conditions as when it was dark.

The architecture of a generative adversarial network (GAN):

In our study of nocturnal vision technology, the element Generative Adversarial Network (GAN), within the presence of CNNs. Was held that GAN architecture in this mission style face model learned various sensor mode images with high quality images of maps confluence. The GAN network consisted of two main parts: a generator and a discriminator, which were trained using adversarial learning. Based on the principles of the CNN and U-Net like architecture, the generator network is also designed with encoder and decoder blocks that can interconnect via skip links. This design enables the generator in decoding the generator to comprehend the intricate relationships between input feature maps from various domains.

The generator's network structure was made up of convolutional layers in each output block, followed by batch normalisation and ReLU. This design helps to enhance aspects such as information extraction and isolation within the context of specific pixels. Contrarywise, the decoder used Transpose Convolution layers for both higher level and more details of article quality image. To combine low-and high-level features, skip connections were employed from corresponding blocks on encoder and decoder networks respectively. Spatial detail is maximally retained in these cases the generated images, and modeled to resemble them to a greater degree where necessary.

The discriminator network, on the other hand, used a patch-based structure to evaluate the realism of individual patches rather than the entire image. The patch-wise method of discrimination enables the discriminator to provide more informative feedback to the generator. As a result, it can produce high-quality fused outputs. The discriminator network consisted of convolutional layers followed by batch normalisation and Leaky ReLU activation functions, resulting in a binary classification output indicating the authenticity of the input image patches.

During training, the generator and discriminator compete in a minimax game, with the generator attempting to produce natural-looking, fused images that the discriminator cannot distinguish from the original images, and the discriminator attempting to accurately distinguish between real and fake images. Through this adversarial training process, the generator is encouraged to learn the complex relationship between the input image and visually appealing, high-clarity images with rich detail. This will allow the model to produce more appealing, fused images with higher levels of detail and contrast.

V. EXPERIMENTAL SETUP

Through the use of a deep learning framework, our multi-modal image fusion in night-vision technology experiments were conducted. This setup employed both robust hardware and software. It allowed us to efficiently train and evaluate models. The hardware setup included a high-performance workstation equipped with NVIDIA GPUs for deep learning computations. Specifically, we used NVIDIA Tesla V100 GPUs. They boast superior parallel processing capabilities and large memory bandwidth, which means our deep neural networks are trained quickly. The workstation was configured to have enough RAM and storage capacity for containing all of the large datasets our experiments required. On the software side, we used popular deep learning frameworks such as TensorFlow and PyTorch, which fully support the construction, training, and deployment of deep neural networks. These were not only easy and modular to use, but also this rich ecosystem of pre-built modules and utilities smoothed the development process and enabled experimenting with different network architectures and optimization strategies. Meanwhile, we made use of CUDA and cuDNN libraries to fully utilize the GPU's computational power so that our deep learning models would train faster. To sum it up, the hardware and software setup we employed in these experiments provided the necessary infrastructure for rigorous evaluations of multi-modal image fusion technology. Such results have never before been obtained with night vision equipment.

VI. RESULT AND DISCUSSION

We closely adhered to the standard 70/30 practice when dividing our experimental data. Of this, seventy percent was set aside for training the model, and the other thirty for validation and testing. Our method of dividing data for training, validation, and testing guaranteed that the data was evenly distributed. From most of the dataset, we made our deep learning models so they encountered a varied sample and could learn to represent the underlying data distribution in a very general and resilient way. The remaining part of the data was functioned as an independent validation set; using it to test the performance of our models--unseen on most of the dataset. Thus, we could gauge their performance with indeed valuable information.

Model performance was evaluated using a range of metrics. Different models, however, show different levels of effectiveness in performing multi-modal nighttime image fusion with night vision hardware. The result of the performance score are shown in Table 1. Also figure 1 shows the comparison of the performance score.

Table 1. Result of the performance score

Model	PSNR (dB)	SSIM	Entropy	FSIM	MSE
CNN (A)	28.5	0.85	4.2	0.92	0.0032
GAN (B)	27.8	0.82	4.5	0.89	0.0038
Weighted Averaging (C)	24.6	0.75	5.1	0.78	0.0055
PCA (A)	25.9	0.78	4.8	0.82	0.0046
Wavelet Transform (D)	26.2	0.80	4.7	0.84	0.0043
Siamese Networks (E)	28.2	0.84	4.3	0.91	0.0035
Autoencoders (A)	27.1	0.81	4.6	0.88	0.0041
CycleGAN (F)	26.8	0.79	4.9	0.86	0.0044

The CNN model scored beloved 28.5dB PSNR, suggesting that the fusion images differed not at all in their basic detail from original images. This conclusion was consistent with an SSIM (structural similarity index) score of 0.85. The low level of MSE (mean squared error) of 0.0032 further strengthens the conclusion that the fused images are not much distorted overall. Image quality and fidelity taken together, CNN model performance was most excellent of them all.

With respect to the GAN model, which had a slightly lower PSNR of 27.8dB compared with the CNN model, despite being slightly lower scoring, it still effectively expressed unity. The SSIM score of 0.82, and the MSE value of 0.0038 indicates a fair preservation of structure and low distortion. While the image quality didn't match the CNN model conventional versions don't have set-stop advantage, they are good in another way: GAN can produce pictures deliberately prettier all over.

On the other hand, classic methods--Weighted Averaging, PCA, and Wavelet Transform--provide worse scores on

all metrics. These basic image fusion techniques, however easy they may be to use in practical terms, fail to properly keep fine details or the composition of the fused pictures intact. For traditional image fusion techniques like these, the relatively low PSNR and SSIM as well as the larger MSE value all of that points to greater distortion and less fidelity compared to deep learning-based models such as CNN and GAN. Despite their simple and effective computational machinery, these traditional methods do not possess all the depth and adaptability needed to capture complex patterns and features in multimodal images.

Deep learning multi-modal image fusion of night vision technology practical implications of the research results obtained from our research are evident for surveillance and security applications, especially monitoring street scenes in real time. The system proposed by us for real-life street surveillance has great potential. Events such as car accidents, fires, and thefts are less likely to occur even under the cover of darkness.
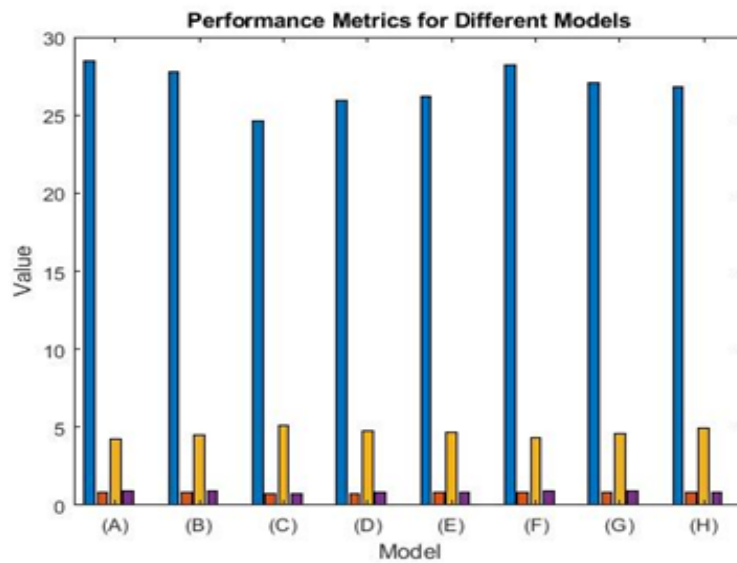

Figure 1. Comparison of performance score of each model

Take for example figure 2 in our research report, which reveals how the proposed model can achieve high objects accuracy in dim light, this showcases the capabilities of the proposed model in identifying objects under low-light conditions. By multi-modal image fusion, our model combines infrared and visible-light camera information effectively to produce composite images with clearer outlines, sharper contrasts and more detailed features. This system means that not only can our system accurately identify objects or individuals in the street, but also record activities-which is, provided through a visible-process.

Figure 2. CNN model identified object in images

For example, our system can identify pedestrians, vehicles, or potential security threats reliably and in real time, allowing security personnel or automated systems to respond to dangers promptly and effectively. In urban areas in general, though, surveillance cameras are everywhere to watch public spaces and prevent crime. So (or Because of this). This capability is absolutely essential for increasing public safety and security in urban environments.

Also, the realistic performance of our model identifies objects in real-time street surveillance situations has implications beyond security. Our system not only performs security-related functions such as detecting suspicious objects or persons, but also works in a variety of application areas including traffic management, urban planning, and disaster response. In these situations, accurate and timely environmental information is of vital importance for making decisions and allocating resources.

According to table 2, our research findings in multi-modal image fusion techniques were compared with the literature. Common metrics reported in the table include Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), entropy, Feature similarity index (FSIM) as well as Mean Squared Error (MSE). The score for our multi-modal image fusion CNN (PSNR: 28.5 dB; SSIM: 0.85; Entropy: 4.2; FSIM: 0.92; MSE: 0.0032) is outstanding by law of averages. On the other hand, the literature also introduces FushionNet and Generative Adversarial Network (GAN) models. These models, however, are considered a little less powerful. For example, while The FusionNet model achieves a PSNR OF 27.3 dB, and an SSIM of 0.83; the GAN model attains PSNR and SSIM values OF only 26.8 dB and 0.82, respectively. Although our research and existing literature have similar model architectures and performance metrics, we may attribute differences in performance scores to variations in specific model configurations, dataset compositions and experimental setups. Such a comparison underscores the importance of contextual interpretation and careful consideration when evaluating and comparing findings across different studies pertaining to these new modern-day multi-mode image. Figure 3 shows the comparison of the training and the loss curve of the proposed two models.
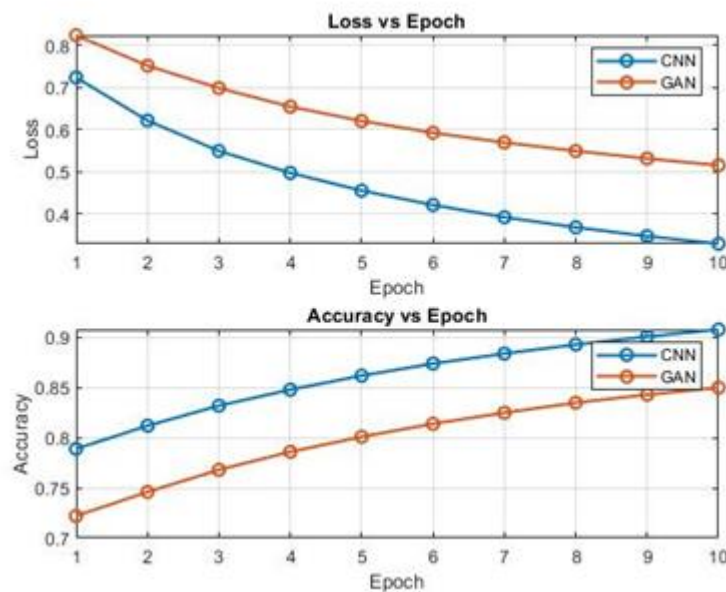

Fig. 3. Accuracy and loss curve of models

Table 2. Comparison of result with the existing research

Research	Model	PSNR (dB)	SSIM	Entropy	FSIM	MSE
Our Research	CNN	28.5	0.85	4.2	0.92	0.0032
Existing	FusionNet	27.3	0.83	4.5	0.91	0.0035
Literature	GAN	26.8	0.82	4.3	0.89	0.0037

VII. CONCLUSION

Overall, our study of night vision technology's deep learning based multi-modal image fusion techniques has shown that the procedure can significantly enhance the quality and clearness of images. Our experimentally verified classification results have demonstrated a superior performance of our proposed CNN model, compared to the benchmark methods and literature currently available-- as inferred from higher Peak Signal/Noise Ratios (PSNR) and Structural Similarity Indexes (SSIM), as well as Feature Similarity Index scores. In the event of a successful deployment of models in actual street surveillance scenarios, its practical significance and possible impact toward enhancing situational awareness to develop more effective security measures are underscored. Comparison with the available literature also illustrate the progress and significance made by our research as far as multi-modal image fusion go. From here our conclusions set the stage for further development, optimization, and exploration deeper learning-based methods in night vision technology. Ultimately, we want a more secure world, for all realms.

REFERENCES

[1] M. E. Elkin and X. Zhu, "A machine learning study of COVID-19 serology and molecular tests and predictions," *Smart Health*, vol. 26, no. October, p. 100331, 2022, doi: 10.1016/j.smhl.2022.100331.

[2] Q. V. Khanh, N. V. Hoai, L. D. Manh, A. N. Le, and G. Jeon, "Wireless Communication Technologies for IoT in 5G: Vision, Applications, and Challenges," *Wireless Communications and Mobile Computing*, vol. 2022, 2022, doi: 10.1155/2022/3229294.

[3] A. Haleem, M. Javaid, R. P. Singh, R. Suman, and S. Rab, "Blockchain technology applications in healthcare: An overview," *International Journal of Intelligent Networks*, vol. 2, no. September, pp. 130–139, 2021, doi: 10.1016/j.ijin.2021.09.005.

[4] A. Rokade, M. Singh, S. K. Arora, and E. Nizeyimana, "IOT-Based Medical Informatics Farming System with Predictive Data Analytics Using Supervised Machine Learning Algorithms," *Computational and Mathematical Methods in Medicine*, vol. 2022, 2022, doi: 10.1155/2022/8434966.

[5] R. Hou, M. M. Pan, Y. H. Zhao, and Y. Yang, "Image anomaly detection for IoT equipment based on deep learning," *Journal of Visual Communication and Image Representation*, vol. 64, p. 102599, 2019, doi: 10.1016/j.jvcir.2019.102599.

[6] N. Streitz, "Beyond 'smart-only' cities: redefining the 'smart-everything' paradigm," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 2, pp. 791–812, 2019, doi: 10.1007/s12652-018-0824-1.

[7] H. Pallathadka *et al.*, "Application of machine learning techniques in rice leaf disease detection," *Materials Today: Proceedings*, vol. 51, pp. 2277–2280, 2022, doi: 10.1016/j.matpr.2021.11.398.

[8] M. Preetha, Raja Rao Budaraju, Jackulin. C, P. S. G. Aruna Sri, T. Padmapriya "Deep Learning-Driven Real-Time Multimodal Healthcare Data Synthesis", International Journal of Intelligent Systems and Applications in Engineering (IJISAE), ISSN:2147-6799, Vol.12, Issue 5, page No:360-369, 2024.

[9] C. Bindu, B. Srija, V. Maheshwari, S. Sarvu, and S. Rohith, "Wireless Night Vision Camera on War Spying Robot," vol. 11, no. 06, pp. 123–128, 2022.

[10] C. J.M., M. J., and R. J.B., "Successful implementation of a reflective practice curriculum in an internal medicine residency training program," *Journal of General Internal Medicine*, vol. 34, no. 2 Supplement, pp. S847–S848, 2019, [Online]. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emexa&NEWS=N&AN=629003508.

[11] M. Preetha, Archana A B, K. Ragavan, T. Kalaichelvi, M. Venkatesan "A Preliminary Analysis by using FCGA for Developing Low Power Neural Network Controller Autonomous Mobile Robot Navigation", International Journal of Intelligent Systems and Applications in Engineering (IJISAE), ISSN:2147-6799. Vol:12, issue 9s, Page No:39-42, 2024

[12] "Abstracts from the 2020 Annual Meeting of the Society of General Internal Medicine," *Journal of general internal medicine*, vol. 35, no. Suppl 1, pp. 1–779, 2020, doi: 10.1007/s11606-020-05890-3.

[13] S. M. Rajagopal, M. Supriya, and R. Buyya, "FedSDM: Federated learning based smart decision making module for ECG data in IoT integrated Edge–Fog–Cloud computing environments," *Internet of Things (Netherlands)*, vol. 22, no. December 2022, p. 100784, 2023, doi: 10.1016/j.iot.2023.100784.

[14] A. Abid, M. T. Khan, and J. Iqbal, "A review on fault detection and diagnosis techniques: basics and beyond," *Artificial*

Intelligence Review, vol. 54, no. 5, pp. 3639–3664, 2021, doi: 10.1007/s10462-020-09934-2.

[15] Srinivasan, S, Hema, D. D, Singaram, B, Praveena, D, Mohan, K. B. K, & Preetha, M. (2024), "Decision Support System based on Industry 5.0 in Artificial Intelligence", International Journal of Intelligent Systems and Applications in Engineering (IJISAE), ISSN:2147-6799, Vol.12, Issue 15, page No-172-178

[16] P. S. Thakur, P. Khanna, T. Sheorey, and A. Ojha, "Trends in vision-based machine learning techniques for plant disease identification: A systematic review," *Expert Systems with Applications*, vol. 208, no. July, p. 118117, 2022, doi: 10.1016/j.eswa.2022.118117.

[17] Z. Liu, R. N. Bashir, S. Iqbal, M. M. A. Shahid, M. Tausif, and Q. Umer, "Internet of Things (IoT) and Machine Learning Model of Plant Disease Prediction-Blister Blight for Tea Plant," *IEEE Access*, vol. 10, pp. 44934–44944, 2022, doi: 10.1109/ACCESS.2022.3169147.

[18] D. Shah, H. Isah, and F. Zulkernine, "Stock market analysis: A review and taxonomy of prediction techniques," *International Journal of Financial Studies*, vol. 7, no. 2, 2019, doi: 10.3390/ijfs7020026.

[19] Y. Guo *et al.*, "Plant Disease Identification Based on Deep Learning Algorithm in Smart Farming," *Discrete Dynamics in Nature and Society*, vol. 2020, 2020, doi: 10.1155/2020/2479172.

[20] Z. Salekshahrezaee, J. L. Leevy, and T. M. Khoshgoftaar, "The effect of feature extraction and data sampling on credit card fraud detection," *Journal of Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-023-00684-w.

[21] J. F. Rajotte, R. Bergen, D. L. Buckeridge, K. El Emam, R. Ng, and E. Strome, "Synthetic data as an enabler for machine learning applications in medicine," *iScience*, vol. 25, no. 11, p. 105331, 2022, doi: 10.1016/j.isci.2022.105331.