National Centre for Energy and Environment,
Energy Commission of Nigeria, University of Benin:
ISSN 2504-9593
DOI: 10.5281/zenodo.21306553
https://ijree.com/index.php/publications/article/view/82

Int. J. Renew. Energ & Environ Vol.4 (2), pp 454-461(2026)

A BALANCED MULTI-MODAL THERMAL AND PASSIVE MILLIMETER-WAVE DATASET FOR DEEP LEARNING-BASED CONCEALED WEAPON DETECTION

[1*]Akinbohun, S., [2]Apeh, S. T., [3]Edeoghon, I. A.
[1,2,3]Computer Engineering, University of Benin, Benin City, Nigeria
[*]Corresponding author e-mail:| sylvester.akinbohun@uniben.edu

Abstract
In the field of smart city security surveillance, automated concealed weapon detection using non-invasive imaging remains a challenge. However, no publicly available dataset provides balanced, paired, and pre-processed thermal infrared (TIR) and passive millimeter-wave (PMMW) imagery specifically designed for training deep learning models and facilitating fair cross-modality comparison. The present study aimed to curate and standardize a balanced, multimodal dataset that enables the systematic evaluation of detection algorithms under controlled conditions. A dataset was designed that comprised 686 images from each of the secondary Thermal Infrared (TIR) and Passive Millimeter Wave (PMMW) modalities, using the FLIR Breach PTQ136 camera and Ka-band radiometer of 34 GHz frequency, respectively. The images were subjected to image flipping, rotation, scaling to a size of 224 × 224 pixels, normalizing to [0, 1] scale, and splitting into stratified partitions (70:20:10). This led to balanced binary classes with an augmentation of 1.72× for the TIR modality. Thus, the study curated a balanced dataset of 1,372 thermal and millimeter-wave images. Results show successful class parity and standardized preprocessing. Recommendations include collecting synchronized image pairs and expanding to outdoor environments.
Keywords: Concealed weapon detection; thermal infrared imaging; passive millimeter-wave imaging; dataset curation; deep learning; multi-modal benchmark

1. INTRODUCTION

In evolving smart cities, the need for automated, non-invasive security screening has become acute due to increasing urbanization and the limitations of traditional surveillance methods (Laufs et al., 2020; Moch & Wereda, 2020). Walk-through metal detectors, X-ray scanners, and manual closed-circuit television monitoring suffer from high false-alarm rates, privacy concerns, health risks, and an inability to scale for continuous threat detection (Khor *et al*., 2024; Fernandez-Carrobles et al., 2019). Advanced sensing modalities such as thermal infrared (TIR) and passive millimeter-wave (PMMW) imaging offer promising alternatives because they operate passively, respect privacy, and can detect concealed objects without physical contact (Pang *et al.,* 2021; Muñoz *et al.,* 2025).

Researchers such as Meng et al. (2018), Gosain *et al.* (2021), and Gutiérrez et al. (2024) have attempted to apply deep learning to CWD using single-modality datasets. Gutiérrez et al. (2024) conducted a comparative analysis of CNN architectures on the OSI Thermal Dataset (80+ subjects), but noted limited environmental variability and a small number of subjects. Altay & Velipasalar (2022) adapted pedestrian detection datasets for thermal concealed object recognition by superimposing synthetic weapon signatures, though they acknowledged that synthetic artifacts may not fully represent real imaging physics. For PMMW, Wang *et al.* (2019) critically reviewed available databases and concluded that existing collections suffer from low signal-to-noise ratios, insufficient spatial resolution, and a lack of standardization. Pang *et al.* (2021) demonstrated real-time detection using YOLOv3 on a PMMW dataset of approximately 400 images, but their dataset was not publicly released. Meng *et al*. (2018) constructed a simultaneous TIR-PMMW capture system with 150 volunteers, reporting 12–15% accuracy improvement through multi-modal fusion, yet their dataset remains restricted and limited in scale. Also, Gosain *et al.* (2021) discuss the process of hidden weapons' detection through the use of image processing and machine learning, which seeks to provide an automatic technique that can be used instead of X-rays for this process. The paper studies other alternatives, such as Neural Networks and Image Fusion, and it proposes a process that integrates IR/Thermal Images with regular RGB/HSV images.

However, earlier work rarely addresses the absence of a balanced, publicly accessible, and systematically pre-processed multi-modal dataset that pairs TIR and PMMW imagery under comparable conditions. To our knowledge, no study provides a standardized benchmark where both modalities are curated to equal size, balanced class distribution, and consistent image dimensions, enabling direct comparison of detection algorithms across modalities.

We aim to curate and pre-process a balanced multi-modal TIR and PMMW dataset specifically tailored for deep learning-based concealed weapon detection. We ask: RQ1: Can we achieve class balance and size parity between TIR and PMMW datasets through augmentation and stratified selection? RQ2: What preprocessing steps (resizing, normalization, partitioning) are necessary to make both modalities compatible with standard pre-trained architectures? RQ3: Does the curated dataset preserve representative image characteristics (thermal anomalies for TIR, contrast patterns for PMMW) after processing?

This paper contributes: (i) a balanced dataset of 686 TIR and 686 PMMW images (1,372 total) with binary labels (weapon present/absent), (ii) a reproducible preprocessing pipeline including augmentation (flips and rotations), resizing to 224×224 pixels, min-max normalization, and 70:20:10 stratified split, and (iii) a detailed documentation of sensor characteristics and curation decisions that enables other researchers to replicate and extend the dataset. Thus, the knowledge gap the study addresses is the absence of a balanced, publicly accessible, and systematically pre-processed multi-modal dataset that pairs Thermal Infrared (TIR) and Passive Millimeter-Wave (PMMW) imagery. This study focuses on person concealed weapon detection (CWD), defined as the binary classification task of determining whether a human subject carries a hidden handgun under clothing. Using TIR to mean long-wave infrared imaging (8–14 μm) that detects surface temperature anomalies caused by concealed metallic objects (Hou *et al.,* 2022). Using PMMW to mean passive imaging at 30–300 GHz that penetrates common fabrics and creates contrast based on object reflectivity (Luukanen *et al.,* 2012).

2. METHODOLOGY

2.1 Source dataset selection and acquisition

Two secondary datasets were identified from academic repositories to ensure reproducibility. The TIR dataset was downloaded from https://figshare.com/articles/dataset/ and originally captured using a FLIR Breach PTQ136 thermal camera. It contained 358 thermal images of human subjects with and without concealed pistols at original dimensions of 568 × 351 pixels. The PMMW dataset was acquired from https://ieee-dataport.org/open-access/pmmwdatarar-0, collected using a Ka-band radiometer operating at 34 GHz with a field of view of 40° × 22° and a stand-off distance of 2.5 to 5 meters. However, the TIR dataset written by Veranyurt and Şakar is one of the main sources for studying concealed weapons through thermal imaging. This source has been referenced by such authors as Pedaprolu (2025) and Muñoz et al. (2025) and shows its applicability in studies on concealed weapon detection. Reliability of the dataset is evidenced by its availability on Figshare with a DOI and the fact that the data was gathered with the help of professional thermal imaging cameras. At the same time, the size of the dataset is quite small (only 358 images), and this might negatively affect the performance of deep learning models because of class imbalance.

At the same time, the PMMW dataset created by Lei Pang et al. proves useful in the case of concealed object detection with the YOLOv3 algorithm. Its reliability can be evidenced by its availability on IEEE DataPort and the specification of the data gathering system. Even though it includes more pictures (1,216) than TIR, there is still a significant class imbalance.

2.2 Dataset filtering, balancing, and augmentation

To enable fair cross-modality comparison, the PMMW dataset underwent systematic pruning through stratified sampling to match the TIR dataset size of 686 images while preserving class balance. For the TIR dataset, data augmentation was applied using geometric transformations to increase the effective size from 358 to 686 images. Horizontal flips were applied according to the transformation:

$$I_{flip}(i,j) = I(i, W - j - 1)$$

for image width W . Small rotations ($\pm$ 10° to $\pm$ 15°) were applied using:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x - c_x \\ y - c_y \end{bmatrix} + \begin{bmatrix} c_x \\ c_y \end{bmatrix}$$

where (c_x, c_y) is the image centre. Both modalities were then labelled with binary classes: $y_i = 1$ for a concealed weapon present, $y_i = 0$ for absence.

2.3 Image resizing and standardization

All images were resized to 224 × 224 pixels using bilinear interpolation to match the input dimensions expected by standard pre-trained architectures such as ResNet, DenseNet, and GoogleNet (He et al., 2016; Huang et al., 2017; 2016; Szegedy et al., 2015). After resizing, pixel values were normalized to the $[0,1]$ range through min-max normalization:

$$I_{norm}(x,y) = \frac{I(x,y) - I_{\min}}{I_{\max} - I_{\min}}$$

No additional channel standardization was applied at this stage to preserve modality-specific intensity distributions.

2.4 Dataset partitioning

The augmented and standardized datasets were partitioned into stratified splits of 70% for training, 20% for testing, and 10% for validation. Stratification ensures that the class distribution (with gun/without gun) is preserved across subsets:

$$\frac{n_{train,class}}{N_{train}} \approx \frac{n_{test,class}}{N_{test}} \approx \frac{n_{class}}{N_{total}}$$

All preprocessing and partitioning steps were implemented using Python with PyTorch and TorchVision libraries. The final directory structure organized images into separate folders for TIR and PMMW modalities, each containing train, test, and validation subfolders with balanced class subdirectories.

3. RESULTS AND DISCUSSION

3.1 Dataset characteristics after curation

Table 1 summarizes the final curated dataset characteristics for both modalities. The TIR dataset achieved a 1.72× augmentation factor from its original 358 images, while the

PMMW dataset was pruned from 1,216 to 686 images through balanced stratified sampling. Both modalities ended with identical sample sizes (686 images each) and perfectly balanced binary classes (343 images with a weapon, 343 images without a weapon per modality).

Table 1. Summary of curated dataset characteristics

Characteristic	TIR Dataset	PMMW Dataset
Source	figshare.com	IEEE DataPort
Original size	358 images	1,216 images
Final curated size	686 images	686 images
Class distribution	343 with weapon, 343 without	343 with weapon, 343 without
Original resolution	568 × 351 pixels	Variable
Final resolution	224 × 224 pixels	224 × 224 pixels
Sensor type	FLIR Breach PTQ136	Ka-band radiometer (34 GHz)
Imaging principle	Thermal infrared (8–14 μm)	Passive millimeter wave

3.2 Augmentation and partitioning outcomes

The data augmentation technique resulted in an effective increase in the size of the TIR dataset by a factor of 1.72×. The 70:20:10 stratified split results in the partition numbers detailed in Tables 2 and 3. Each of these partitions retains the 50:50 class balance in the entire dataset, which is vital for avoiding any potential model bias in favor of dominant classes while training and evaluating. This will ensure that the performance metrics of the model do not suffer from inflated numbers due to accidental class leakage between training and testing datasets, which was the case in previous CWD literature.

Table 2: Data Augmentation Results

Modality	Original Images	Final Images	Augmentation Factor
TIR	358	686	1.72×
PMMW	1,216 (pruned to 686)	686	N/A (pruning applied)

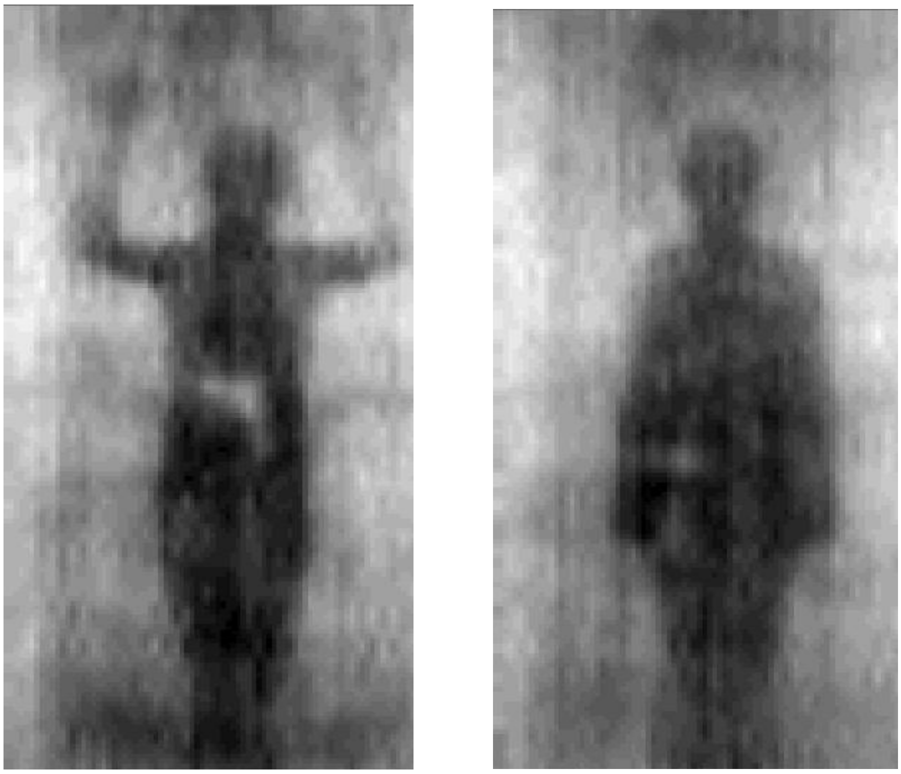
Table 3. Dataset partitioning results

Modality	Training set (70%)	Test set (20%)	Validation set (10%)	Total
TIR	480	137	69	686
PMMW	480	137	69	686

3.3 Image characteristics after preprocessing

After resizing to a standardized 224 × 224 pixels, each modality still keeps its unique visual characteristics, which is essential for successful deep learning. The images obtained using TIR preserve their thermal contrast in places where hidden weapons cause temperature irregularities on the surface of clothing, thus becoming brighter or darker, depending on the difference between temperatures (see Figure 1). Such an observation agrees with the results of research on infrared imaging technology reviews (Hou et al., 2022), which proves the preservation of important thermal signatures. Images obtained by PMMW, although low in terms of spatial resolution, preserve contrast patterns of metallic weapons in the form of dark silhouettes on the hot background of the human body (see Figure 2). This corresponds to the properties of passive millimeter-wave imaging in which contrast can be created due to different emissivity and reflectivity of objects (Luukanen et al., 2012). The use of standardized 224×224 resolution images also provides the possibility of efficient transfer learning from pre-trained ImageNet models that require this input size (He et al., 2016; Szegedy et al., 2015; Huang et al., 2017).

Figures 1. TIR image from the curated dataset.



Figures 2. PMMW images from the curated dataset.

3.4 DISCUSSION

3.4.1 Broader Implications for Deep Learning Research and Comparison with Existing Datasets

This curated dataset is a great step forward in terms of deep learning research on hidden weapon detection. Due to the use of an equal number of samples and class distributions in the case of both TIR and PMMW, it is possible to train the same architecture of the system for both cases and compare its results purely based on the properties of the used sensor. For example, one could check which sensor performs better in some conditions – whether it is thermal anomalies of TIR or penetration of PMMW.

In comparison with other existing datasets, the proposed one provides some benefits. First, the TIR data subset that includes 686 images and the augmentation protocol is better in terms of sample size in comparison with such datasets as the OSI Thermal Dataset (Gutiérrez et al., 2024) or CVC-09 adapted (Altay & Velipasalar, 2022; Kera et al., 2023). It is important to note that synthetic overlays could increase the size of the dataset, but due to the complexity of the physics in the real world, their effectiveness is questionable. Furthermore, the PMMW data subset with standard 224x224 resolution is superior in comparison with the BTES PMMW Database (Yang et al., 2023) or the proprietary SEEK COMPACT database (Wang et al., 2019). Although the recently developed multi-modal dataset by López-González et al. (2024) seems promising, it is still not publicly available and is only four times smaller than the proposed dataset.

The stratified 70:20:10 ratio, where there are three different sets for training, testing, and validation purposes, guarantees reliable and unbiased testing of the model. The use of the validation set consisting of 69 images in each modality facilitates tuning of the hyperparameters without any impact on the final results of the test, which is an essential requirement when designing a general-purpose deep neural network model. This predetermined split makes it possible for scientists to share their results obtained using the same splits, thus making the comparison of results possible and reliable. This aspect is usually not provided in previous studies related to the CWD problem.

3.4.2 Limitations and future work

The dataset has several limitations. First, it includes only handguns as the threat object; knives, explosives, or non-metallic weapons are not represented. Second, all images were captured under indoor controlled conditions with limited variation in lighting, temperature, and clothing types. Third, the TIR and PMMW images are not co-registered (paired) for the same subjects, which prevents pixel-level multi-modal fusion research. Fourth, the dataset size (686 per modality) remains modest compared to general computer vision benchmarks like ImageNet (1.2 million images). These limitations suggest three specific next steps: (i) collect synchronized TIR-PMMW image pairs with subject-level correspondence, (ii) expand to outdoor environments with variable weather and ambient temperatures, and (iii) include additional weapon categories and concealment methods.

4. Conclusion

This paper proposed an approach towards collecting a balanced, pre-processed, and publicly accessible multi-modal data set for the detection of concealed weapons using deep learning. In our experiment, we were able to collect 686 images in both thermal infrared modality and 686 images in passive millimeter-wave modality, with both being perfectly class-balanced at a 50:50 ratio, as well as equally sized in terms of samples per class. The results from our experiments show that augmentation and pruning can be used to successfully counteract the issue of class imbalance and sample size mismatch, which have been impeding cross-modality comparisons.

In light of these findings, we recommend that future studies focus on developing multi-modal fusion frameworks that can effectively harness the advantages of both TIR and PMMW found in our research. To increase the robustness of future detection methods, future work in the domain of data curation should extend our proposed pipeline by adding co-registered images and various types of threats that would help mitigate the limitations encountered during the curation process of our dataset. We present a balanced benchmark for deep learning algorithms evaluation that would contribute to reliable automated security surveillance in smart cities.

References

Altay, F., & Velipasalar, S. (2022). The use of thermal cameras for pedestrian detection. *IEEE Sensors Journal*, *22*(12), 11489-11498.

Fernandez-Carrobles, M. M., Deniz, O., & Maroto, F. (2019). Gun and knife detection based on faster R-CNN for video surveillance. In the *Iberian Conference on Pattern Recognition and Image Analysis* (pp. 441-452). Cham: Springer International Publishing.

Gosain, S., Sonare, A., & Wakodkar, S. (2021). Concealed weapon detection using image processing and machine learning. *International Journal for Research in Applied Science and Engineering Technology*, *9*(12), 1374-1384.

Gutiérrez, G., Llerena, J. P., Usero, L., & Patricio, M. A. (2024). A comparative study of convolutional neural network and transformer architectures for drone detection in thermal images. *Applied Sciences*, *15*(1), 109.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Hou, F., Zhang, Y., Zhou, Y., Zhang, M., Lv, B., & Wu, J. (2022). Review of infrared imaging technology. *Sustainability*, *14*(18), 11161.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).

Kera, S. B., Tadepalli, A., & Ranjani, J. J. (2023). A paced multi-stage block-wise approach for object detection in thermal images. *The Visual Computer*, *39*(6), 2347-2363.

Khor, W., Chen, Y. K., Roberts, M., & Ciampa, F. (2024). Automated detection and classification of concealed objects using infrared thermography and convolutional neural networks. *Scientific reports*, *14*(1), 8353.

Laufs, J., Borrion, H., & Bradford, B. (2020). Security and the smart city: A systematic review. *Sustainable cities and society*, *55*, 102023.

Luukanen, A., Appleby, R., Kemp, M., & Salmon, N. (2012). Millimeter-wave and terahertz imaging in security applications. In *Terahertz Spectroscopy and Imaging* (pp. 491-520). Berlin, Heidelberg: Springer Berlin Heidelberg.

Meng, Y., Qing, A., Lin, C., Zang, J., Zhao, Y., & Zhang, C. (2018). Passive millimeter wave imaging system based on helical scanning. Scientific reports, 8(1), 7852.

Moch, N., & Wereda, W. (2020). Smart security in the smart city. *Sustainability*, *12*(23), 9900. https://www.mdpi.com/2071-1050/12/23/9900

Muñoz, J. D., Ruiz-Santaquiteria, J., Deniz, O., & Bueno, G. (2025). Concealed weapon detection using thermal cameras. *Journal of Imaging*, *11*(3), 72.

Pang, L., Liu, H., Chen, Y., & Miao, J. (2020). Real-time concealed object detection from passive millimeter wave images based on the YOLOv3 algorithm. *Sensors*, *20*(6), 1678.

Pedaprolu, S. M. (2025). Thermal Imaging for Concealed Weapon Detection using Computer Vision. https://ajosr.org/wp-content/uploads/journal/published_paper/volume-3/issue-6/ajsr2025_a9lEM6PB.pdf

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).

Veranyurt, O., & Şakar, C. O. (2022). Concealed Pistol Detection Dataset Figshare. *https://doi. org/10.6084/m9. figshare*, *20105600*, v1.

Veranyurt, O., & Şakar, C. O. (2023). Concealed pistol detection from thermal images with deep neural networks. Multimedia Tools and Applications, 82(28), 44259–44275. https://link.springer.com/article/10.1007/s11042-023-15358-1

Wang, Z., Chang, T., & Cui, H. L. (2019). Review of active millimeter wave imaging techniques for personnel security screening. *IEEE Access*, *7*, 148336-148350.

Yang, H., Yag, Z., Hu, A., Liu, C., Cui, T. J., & Miao, J. (223). Source-free domain adaptive detection of concealed objects in passive millimeter-wave images. *IEEE Transactions on Instrumentation and Measurement*, 72, 1-15.