This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3530900

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

Diagnosing Grass Seed Infestation: Convolutional Neural Network Based Terahertz Imaging

QIGEJIAN WANG[1] †, (Graduate Student Member, IEEE), AMUS CHEE YUEN GOAY[1] †, (Student Member, IEEE), DEEPAK MISHRA[1], (Senior Member, IEEE), EWA M. GOLDYS[2], and SHAGHIK ATAKARAMIANS[1], (Senior Member, IEEE)

[1]School of Electrical Engineering and Telecommunications, University of New South Wales (UNSW), Sydney, NSW 2052, Australia
[2]School of Biomedical Engineering, University of New South Wales (UNSW), Sydney, NSW 2052, Australia
† These authors contributed equally to this work.

Corresponding author: Shaghik Atakaramians (e-mail: s.atakaramians@unsw.edu.au).

This work was supported by Keysight Technologies.

ABSTRACT Grass seed infestation is a significant issue in the Australian sheep industry. Detecting the seeds when they are in wool or on the surface of the skin could assist with prevention of the grass seed infestation. Terahertz imaging provides a viable option for detecting seeds due to its short wavelength, non-ionizing feature, and penetration ability through wool. Here we demonstrate that accuracy of seeds detection can be improved utilising a Convolutional Neural Network even when the seeds are not visually distinguishable in terahertz images. Our studies reveal accuracies of greater than 95% and 67% can be achieved in identification of seed hidden underneath 1 cm and 2 cm thick wool under normal incidence. Moreover, our analysis finds that terahertz frequencies in the 0.3–0.4 THz range have better overall classification accuracy compared to other frequency bands. The combination of machine learning and terahertz imaging has the potential to be widely implemented in rapid and on-site detection of grass seed infestation with high efficiency.

INDEX TERMS Convolutional Neural Network, Grass seed infestation, Terahertz imaging.

I. INTRODUCTION

GRASS seed infestation (GSI) has been a significant issue in the Australian sheep industry. The seeds are picked up in the fleece when sheep are grazing. If not acted upon in a timely manner, the seeds may penetrate through the pelt in a few days, leading to contamination and skin damage. Problems caused by GSI include infections and feet inflammation, lower growth rate, and carcasses with less volume, leading to low quality meat and wool [1], [2]. The losses due to GSI of the Australian market alone is about AUD 47.5 million yearly, highlighting the magnitude of the problem [3]. As a result, some seed management strategies have been applied by Meat and Livestock Australia including grazing management, limiting the size of grazing grounds, premature shearing, and using herbicides on the grass early in the growing season [4]. These management strategies come with a trade-off and high cost, which are impractical in the long run for addressing the GSI problem. Therefore, early surface detection of seeds before they penetrate the skin is potentially helpful for GSI management.

Imaging is an option that can be used for prevention or diagnosis of GSI. However, commercial imaging technologies such as computed tomography (CT), ultrasonography, fluorescent imaging, and infrared (IR) imaging are not feasible for seed detection. Note that CT imaging uses ionizing X-rays and ultrasonography requires skin contact, which may cause health issues in the livestock [5], [6]. Fluorescent and IR radiation cannot penetrate through wool due to high scattering effects, and thus are not able to identify seeds [7].

The terahertz band refers to the region of the electromagnetic spectrum with frequency at 0.1–10 THz (3000–30 μm of wavelength) [8]. The non-ionizing nature of the terahertz radiation has made it attractive for health and biomedical applications, which is ideal for non-contact measurements and does not cause any harm to biological species [9], [10]. Therefore, terahertz spectroscopy has potential to be used to detect seeds in sheep fleece, contributing to the prevention of GSI. In our previous work, we demonstrated that terahertz waves can be utilized for the early detection of seeds in the animal fleece or on the pelt [11], [12]. Our experimental results show that terahertz imaging can clearly detect seeds in uncovered ham, pork skin and wool, which can be easily observed from RGB images by the human eye. But when concealed by a thick layer of wool, which is the same scenario as grass seed in-

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3530900

festation, it becomes challenging to discern seeds from RGB images by the human eye. Moreover, human eye observation has limitations in terms of identification speed and accuracy, which is time-consuming and inefficient for batch image processing. Hence, machine learning techniques can assist with image processing, particularly for tasks such as identification and classification [13]–[16]. Various approaches utilizing machine learning techniques for terahertz image processing have been investigated. For instance, a Support Vector Machine (SVM) with parameter optimization has effectively classified transgenic and bacterial blight resistant seeds [17]–[19]. Principal Component Analysis (PCA) followed by SVM were applied to identify moldy wheat, achieving a prediction accuracy of over 90% [20]. Furthermore, SVM and Deep Neural Network (DNN) methods achieve high accuracies in classifying glucose and lactose frequency spectra, providing classification accuracies of 99% and 89.6%, respectively [21]. Convolutional Neural Networks (CNNs) have proven to have improved accuracy over other widely used machine learning methods in real-time liquid contraband classification [22], as well as in the detection of bacterial blight-resistant rice seeds [19].

Inspired by existing works, CNNs are potentially an ideal option for improving GSI diagnosis. By contrast with other applications where machine learning has been used for terahertz imaging, in this work the seeds are covered by wool, making it more difficult to detect due to the scattering and blocking of the terahertz beam. Whereas in most of other applications there is no blockage between the samples and terahertz beam. In addition, unlike most of the reported works that have used flat containers to hold the seeds, for GSI the seeds are usually on or in the skin, which is not flat and may cause more scattering compared to a flat surface. Moreover, GSI may have different type of seeds with various shapes and orientations, which will affect the accuracy. Previously, we have demonstrated that one cumin seed on a piece of ham covered by 1 cm-thick wool can be detected using CNNs, providing an identification accuracy of 95.8% [23].

In this work, we combine machine learning with terahertz imaging to improve accuracy of seed identification. Terahertz images are acquired using time-domain pixel-by-pixel reflection scans with normal and 45° incidences. The samples include three types of seeds placed on a piece of ham, which are covered by wool with different thickness. We investigate the detection accuracy for different terahertz frequencies. The effect of wool thickness, seed shape and orientation are also studied. Additionally, we also compare the results with normal and 45° incident beams. Under normal incidence, the detecting accuracies of $> 95\%$ and $> 67\%$ are achieved when the sample is covered by 1 cm and 2 cm wool, respectively. Whereas under 45° incidence the detecting accuracy is $> 71\%$ when covered by 1 cm wool.

II. TERAHERTZ IMAGE COLLECTION

Generally, there are two types of experimental setup for terahertz imaging including transmission mode and reflection mode [24]. In transmission mode, the terahertz wave needs to penetrate the sample [25], [26], whereas in reflection mode, the terahertz wave is reflected by the sample [27], [28]. Moreover, reflection mode can be divided into oblique reflection and normal reflection. For normal reflection, beam splitting is often needed when the transmitter and receiver are not integrated in one port. Compared to oblique reflection, normal reflection will lose part of the terahertz power due to the beam splitter, but the alignment is much easier. Most of the existing studies on seed detection use transmission or oblique reflection mode [18], [19], [29]–[31]. In this work, since transmission mode is not feasible for seed identification in the pelt, we utilize oblique and normal reflection modes for terahertz imaging.

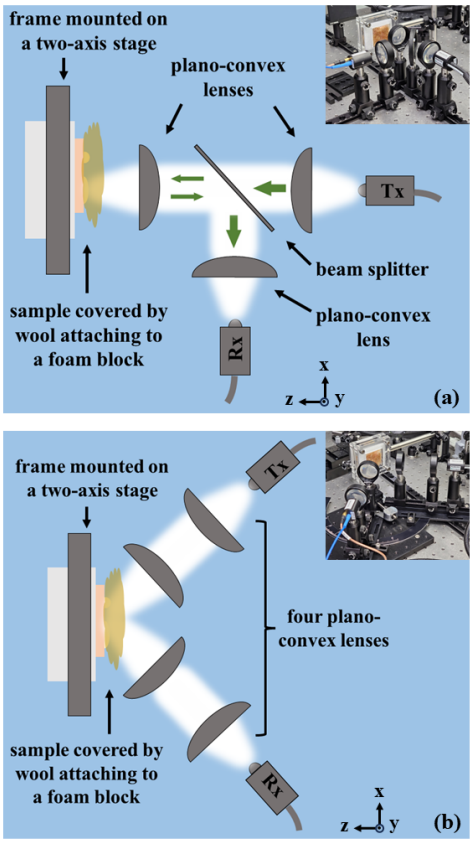


FIGURE 1. Schematic of the experimental setup with (a) normal and (b) 45° incidence. The insets are photos of the experimental setup.

We use a TERASMART time-domain terahertz spectrometer from Menlo Systems, which has a bandwidth > 5 THz and dynamic range > 90 dB. Two modes of experimental setup are used for the measurements including normal (Fig. 1 (a)) and 45° oblique (Fig. 1 (b)) incidence, where the electric field is polarized along the y direction, and the sample is set in a frame mounted on a two-axis stage moving in xy-plane. For normal incidence shown in Fig. 1 (a), we use three polymethylpentene (TPX) lenses to focus the terahertz beam to the sample and detector. A silicon beam splitter is used to split the beam and reflect it to the terahertz detector. For 45°

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3530900

incidence shown in Fig. 1 (b), the emitter and the detector are positioned at 45 degrees from the normal to the surface, and four TPX lenses are used to focus the terahertz beam to the sample and back to the detector. The samples, which are mounted in a frame (see insets of Fig. 1), are scanned pixel by pixel with a step size of 0.5 mm. This process ensures that the terahertz beam is focused on each pixel. Similar setups and imaging methods have been used in our previous work [11].

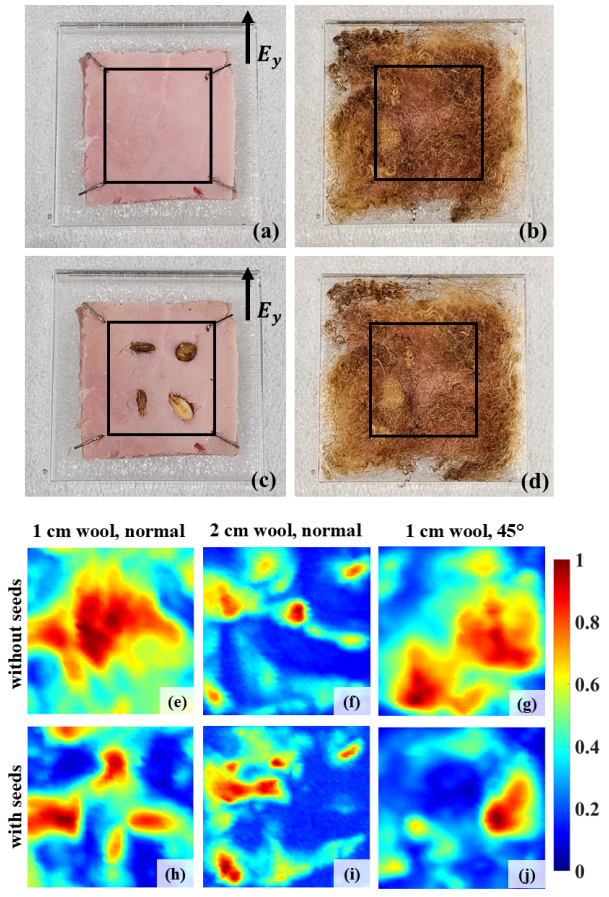


FIGURE 2. 5 mm-thick ham without seeds (a) exposed and (b) covered by 1 cm-thick wool. (c) and (d): Samples with four seeds corresponding to (a) and (b). Normalized time-domain peak-to-peak amplitude of sample in part (a) covered by (e) 1 cm-thick and (f) 2 cm-thick wool under normal incidence. (g) Normalized time-domain peak-to-peak amplitude of sample in part (a) covered by 1 cm-thick wool under 45° incidence. Normalized time-domain peak-to-peak amplitude of sample in part (c) covered by (h) 1 cm-thick and (i) 2 cm-thick wool under normal incidence. (j) Normalized time-domain peak-to-peak amplitude of sample in part (c) covered by 1 cm-thick wool under 45° incidence.

The samples used for experiments are a piece of 5 mm-thick ham (Fig. 2 (a)), which are covered by wool (density 0.04 g/cm^3) with 1 cm or 2 cm thickness when being scanned. For reference, we firstly scan samples without seed (Fig. 2 (b)). Then four seeds are inserted to the same ham sample including two cumin seeds in different orientations, one coriander, and one barley seed (Fig. 2 (c)), which are then covered by wool (Fig. 2 (d)). Since fresh grass seeds contain more moisture than the packed ones, the seeds are soaked in tap water for 3 hours before inserting to the ham. In reality, it is difficult to maintain the surface of the samples to be exactly perpendicular to the ground. Therefore, to include the effect of this factor, each sample is scanned three times with different rotation along x-axis, including zero (perpendicular to the ground) and $\pm 5°$.

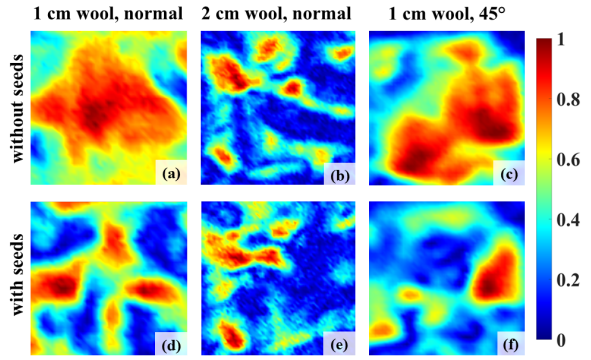


FIGURE 3. (a)–(c) Normalized spectrum amplitude at 0.3 THz corresponding to Figs. 2 (e)–(g). (d)–(f) Normalized spectrum amplitude at 0.3 THz corresponding to Figs. 2 (h)–(j).

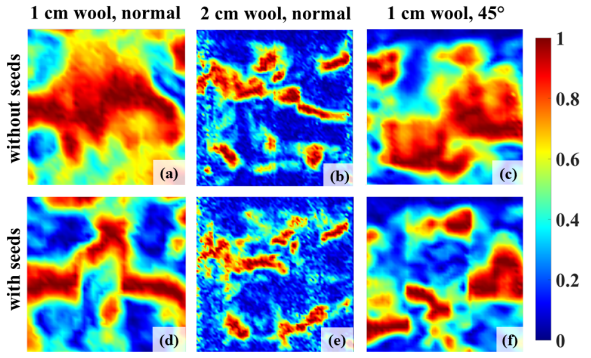


FIGURE 4. (a)–(c) Normalized spectrum amplitude at 0.4 THz corresponding to Figs. 2 (e)–(g). (d)–(f) Normalized spectrum amplitude at 0.4 THz corresponding to Figs. 2 (h)–(j).

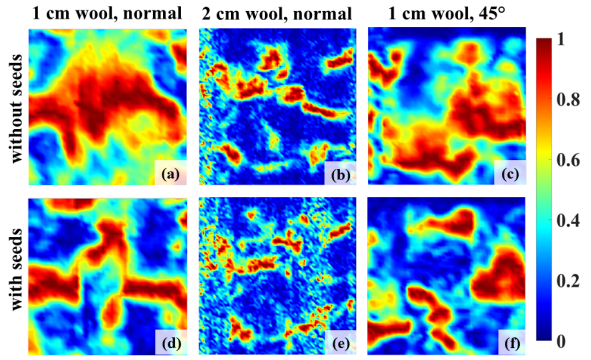


FIGURE 5. (a)–(c) Normalized spectrum amplitude at 0.5 THz corresponding to Figs. 2 (e)–(g). (d)–(f) Normalized spectrum amplitude at 0.5 THz corresponding to Figs. 2 (h)–(j).

Figure 2 (e)–(j) show the normalized time-domain peak-to-peak amplitude, where the data are normalized to the local maximum of each image. Generally, existence of seeds

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3530900

will increase the deep blue area due to the scattering. When covered by 1 cm wool, the area where the seeds are located can be roughly observed by eye (Fig. 2 (h)). However, for samples covered by 2 cm wool (Fig. 2 (i)) or measured using 45° incidence (Fig. 2 (j)), it is difficult to observe by the human eye. To find the optimized frequencies for seed identification, we apply a Fast Fourier Transform (FFT) to the time-domain data and obtain frequency-domain images from 0.2 THz to 0.5 THz with a step of 0.02 THz. Figure 3–5 shows the normalized amplitude at 0.3 THz, 0.4 THz and 0.5 THz, respectively. The seeds are discernible to the human eye when the sample is covered with 1 cm wool and illuminated by normal incidence (Fig. 3 (d), 4 (d) and 5 (d)), whereas identification becomes challenging with 2 cm wool (Fig. 3 (e), 4 (e) and 5 (e)) or 45° incidence (Fig. 3 (f), 4 (f) and 5 (f)). The contrast between samples with and without seeds is less pronounced in terahertz images, with the difference expected to diminish further with increased wool thickness. Consequently, this motivates the application of machine learning techniques to analyze terahertz images for grass seed identification, given the limitations of human eye observation in such scenarios.

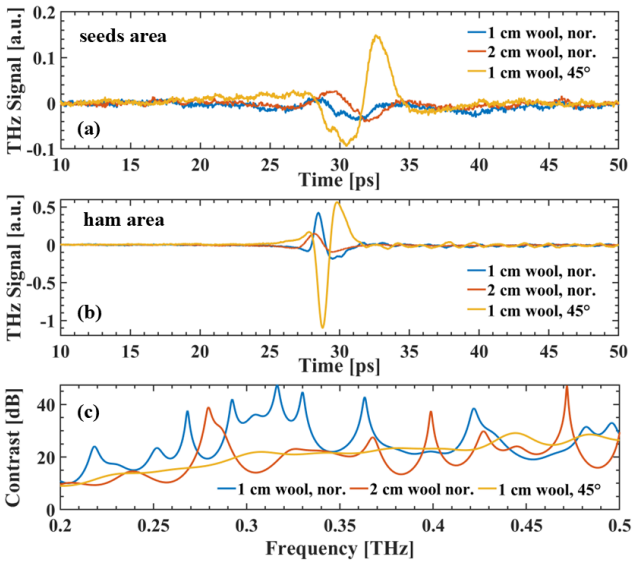


FIGURE 6. Measured time-domain signals of (a) seeds area and (b) ham area covered by 1 cm and 2 cm wool under normal and 45° incidence. (c) Contrast spectrum between ham and seeds area for 1 cm and 2 cm wool under normal and 45° incidence.

Figure 6 (a) and (b) show examples of measured terahertz signals at seeds area (the first cumin seed of each sample) and ham area, respectively. It can be observed that under normal incidence in a seed region, the sample with 1 cm wool has slightly weaker signal than the sample with 2 cm wool, which may be due to the fact that more signal can reach the seed and scatter away, leading to less signal reflected to the detector. Whereas in a ham region, the sample with 1 cm wool has stronger signal than the sample with 2 cm wool due to improved reflection. This is why the sample with 1 cm wool has a higher terahertz contrast in the seed region than sample with 2 cm wool. For a 45° incidence, the reflected signal is higher than at normal incidence in both the seed and ham regions due to the lack of power loss from the beam splitter. Figure 6 (c) shows the contrast spectrum between the ham and seed regions. It can be observed that the sample with 1 cm wool under normal incidence has the overall highest contrast, which is consistent with the observations from Fig. 3–5. It is noteworthy that in this work no reference signal is needed because the wool and seeds only affect the signal strength by attenuating and scattering it. Therefore, by examining the signal strength (weaker when it encounters the seed) and shape for areas with strong contrast, we can identify the existence of foreign matter.

III. TERAHERTZ IMAGE PROCESSING

A. CONVOLUTIONAL NEURAL NETWORK FRAMEWORK

Terahertz imaging is an emerging modality that offers detailed insights into the internal and surface properties of materials, making it a promising tool for this kind of biological inspection. However, due to the complexity of data generated by terahertz imaging, manual analysis becomes impractical. Consequently, we employ CNNs, a class of DNNs specifically designed for image processing tasks, to automate the detection and classification process. Notably, CNNs excel at recognizing patterns in image data by capturing intricate features through a hierarchical, multi-layer structure [32]. The use of convolutional filters that effectively process and extract relevant features prior to classification, makes CNNs particularly well-suited for identifying grass seeds based on the subtle variations present in terahertz images. Additionally, studies have demonstrated that CNNs outperform other widely used machine learning methods on terahertz image classification [19], [22]. Moreover, CNNs have been compared with SVMs in image classification tasks, showing significant improvements in accuracy [33], [34].

The CNN architecture is structured in such a way that it progressively learns to identify increasingly complex features within the images. The initial layers of a CNN typically detect basic patterns, such as edges and textures, while deeper layers capture more abstract and complex features, such as shapes or specific object details. In our case, this hierarchical feature extraction process assists in distinguishing images containing seeds and those without, based on terahertz data.

Since our experiment requires significant time to collect data, we do not have a large enough dataset to construct a CNN from scratch. Therefore, we adopt the transfer learning technique, selecting a well-established, pre-trained CNN based on a large dataset, and retrain it for the seed identification task [35]. This approach significantly simplifies the CNN construction process while enabling us to train our model with our dataset. In this context, we utilize the pre-trained GoogleNet model[1] and fine-tune it with our terahertz image data using MATLAB Deep Network Designer tool.

[1]GoogleNet is a 22-layer deep convolutional neural network, with 144 layers when counting all operations, originally designed for ImageNet classification across 1,000 categories such as animals and everyday objects.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3530900

Specifically, we implement two changes to the GoogleNet network structure. First, we replace the third-to-last layer, the 'fullyConnectedLayer', with a new one, setting the input size to 2, and adjusting the weight and bias learning rate factors to 10. Next, we replace the final 'classificationLayer' to ensure the output size is automatically set to 2, allowing the network to classify the samples as either containing a seed or not.

The CNN model undergoes gradual adjustments to its neuron weights throughout the training process, iteratively refining its ability to classify samples based on the presence or absence of seeds. To facilitate this, we set a low initial learning rate of 0.0001 in the training option, enabling the CNN model to fine-tune its neuron weights using our data. Additionally, we set the MiniBatchSize to 50, MaxEpochs of 100 and ValidationFrequency to 1. All other parameters are left at their default settings to maintain general training behavior.

By employing this approach, we leverage the power of deep learning and transfer learning to address the challenge of seed detection in terahertz images with a limited dataset, while ensuring our CNN model is effectively trained for accurate classification. This method provides a practical solution to the constraints of our experimental setup, offering a balance between model complexity and data availability.

B. RESULTS AND DISCUSSION

1) Impact of frequency band, wool thickness, and incidence scheme

The CNN model's performance is evaluated using classification accuracy, which reflects its ability to correctly identify samples with or without seeds. During training, the model's validation accuracy converges after a certain number of iterations, indicating it has reached an optimized state. Given that data was collected from normal and 45° incidence angles across different wool thicknesses (1 cm and 2 cm), we trained the CNN model separately on each dataset and determined the validation accuracies. This thorough approach yielded varied accuracies across different frequency bands, as summarized in Table 1.

Table 1 shows the impact of wool thicknesses and incidence scheme in different frequency bands, where 80% of the data are used for training and the rest 20% of the data are used for validation. It can be observed that under normal incidence, thicker wool leads to a drop in accuracy across all frequency bands. This decline suggests that the CNN model faced difficulties in accurately classifying seed presence in samples with thicker wool, potentially due to increased attenuation and scattering of the terahertz beam. Notably, the minimum classification accuracy observed for 1 cm wool in the selected frequency bands is 92.9%, which is significantly higher than the 67.9% achieved for 2 cm wool. Furthermore, under the normal incidence with 1 cm wool, both the 0.2–0.5 THz band and its sub-bands achieve a classification accuracy exceeding 90%, meeting the desired performance threshold.

Moreover, upon comparing the accuracy between normal and 45° incidence, we consistently find that normal incidence outperforms the 45° incidence approach with 1 cm wool, as given in Table 1. In particular, the 45° incidence setup achieves a maximum classification accuracy of 85.7% in the frequency band of 0.3–0.4 THz, while the normal incidence setup attains a perfect classification accuracy of 100% in the 0.4–0.5 THz band. One reason is that 45° incidence has longer traveling distance than normal incidence (in this case 2 cm for normal incidence, 2.82 cm for 45° incidence), leading to more scattering and attenuation from the wool. Another factor is the system alignment for 45° incidence is more complicated than normal incidence due to the non-flat surface of the ham sample. Additionally, it is noteworthy that for normal incidence with 2 cm wool and 45° incidence, the 0.3–0.4 THz frequency band yielded the highest validation accuracy. While the 1 cm normal incidence setup achieves very high accuracy across multiple frequency bands, our analysis shows that the 0.3–0.4 THz band consistently delivers good performance across all setups, making it the optimal frequency band for seed detection.

TABLE 1. Validation accuracy with different wool thicknesses for different frequency bands under normal and 45° incidence.

	1 cm wool, nor.	2 cm wool, nor.	1 cm wool, 45°
0.2–0.5 THz	95.4%	67.9%	71.1%
0.2–0.3 THz	92.9%	70.8%	71.4%
0.3–0.4 THz	92.9%	79.2%	85.7%
0.4–0.5 THz	100%	75%	78.6%

2) Impact of seed shape and orientation

We also study the effect of seed shape and orientation. The sample includes two cumin seeds placed horizontally (named seed 1) and vertically (named seed 3), and two coriander seeds (named seed 2 and 4) on a piece of 3-layer ham covered by 2 cm wool (see fig. 7 (a) and (b)), which is scanned using normal incidence. Figure 7 (c)–(h) show the normalized spectrum amplitude of sample without and with seeds covered by 2 cm wool at 0.3 THz, 0.4 THz and 0.5 THz, respectively. As expected, the identification becomes challenging in terahertz images with 2 cm wool. We divide the sample into four sections, each containing one seed. Then we feed the terahertz images at 0.2–0.5 THz band including independent seed and the combinations of same type of seed (seed 1 & 3, seed 2 & 4) to the CNN model for identification. Table 2 shows the validation accuracy for each seed and combination. It can be observed that coriander seeds have higher validation accuracy than cumin seeds for both separated and combined cases. Specifically, our setup achieves 100% accuracy in classifying samples containing coriander seeds in the separated case, while the accuracy drops to 64.3% for Seed 3 (cumin). Even when combining seeds of the same type for classification, the accuracy for detecting coriander seeds in the absence of other seeds remains high at 92.1%, whereas it is lower for cumin seeds at 75%.

We attribute this to the potential excitation of resonances in the coriander seed [12]. Also, the direction of electric field

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3530900

plays an important role on validation accuracy, especially for the cumin seeds with ellipsoid structure, leading to the dramatic difference for seed 1 and seed 3.

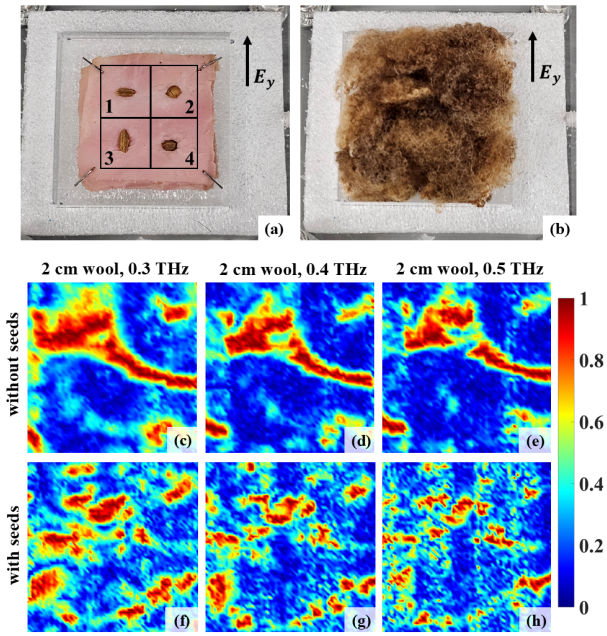


FIGURE 7. Sample (a) exposed and (b) covered by 2 cm wool for investigating impact from different shape and orientation of the seeds, where seed 1 and 3 are cumin, and seeds 2 and 4 are coriander). Normalized spectrum amplitude of sample without seed covered by 2 cm wool at (c) 0.3 THz, (d) 0.4 THz and (e) 0.5 THz. Normalized spectrum amplitude of sample with four seeds covered by 2 cm wool at (f) 0.3 THz, (g) 0.4 THz and (h) 0.5 THz.

TABLE 2. Validation accuracy for different type of seeds with 2 cm wool thickness and normal incidence.

Seed	Accuracy
S1 (cumin)	85.7%
S3 (cumin)	64.3%
S2 (coriander)	100%
S4 (coriander)	100%
S1 & S3 (cumin)	75%
S2 & S4 (coriander)	92.1%

3) Impact of tilted sample

Furthermore, we examine how the accuracy varied concerning different tilted angles ($\pm 5^\circ$ rotation along x-axis, see section II) and wool thicknesses for the terahertz images ranging from 0.2 to 0.5 THz, as summarized in Table 3. The data are from measurements using normal reflection mode. The CNN model demonstrated higher effectiveness in identifying seeds in samples with 1 cm wool compared to those with 2 cm wool. When the sample is angled, we observe higher accuracy, which could be due to the non-perfect alignment and the non-flat surface of the ham. Another reason could be that when positioned in an oblique way, the seeds may have more scattering, leading to higher contrast in the images. Overall, the results show that the $\pm 5^\circ$ rotation along x-axis will not reduce the accuracy.

TABLE 3. Validation accuracy for rotation along x-axis with 1 cm and 2 cm wool thickness under normal incidence.

	1 cm wool	2 cm wool
No rotation	96.2%	69.2%
Vertically tilted 5 degrees	100%	80.8%
Vertically tilted -5 degrees	96.2%	92.3%

C. PERFORMANCE EVALUATION USING CONFUSION MATRICES

To further evaluate the CNN model's performance, we generate confusion matrices that show the classification accuracy by testing the trained model on the entire dataset. Fig. 8 and 10 illustrate the confusion matrices for the CNN model trained with datasets collected through normal and 45° incidence with 1 cm wool thickness across various frequency bands, respectively, whereas Fig. 9 depicts the confusion matrices for normal incidence with 2 cm wool thickness. Upon closer examination of Fig. 8–10, we observe that the majority of errors occur when the CNN misclassified samples without seeds as containing seeds. This could be due to the wool and defects on the ham surface can create same contrast as the seeds do. Moreover, substantial improvement in classification accuracy has been observed for all the frequency bands in all three cases due to the larger dataset and better training, especially for 1 cm wool under 45° incidence. Overall, 1 cm wool under normal incidence still has the highest accuracy, followed by 1 cm wool under 45° incidence. And 2 cm wool under normal incidence still has the lowest accuracy. Focusing on the optimal frequency band, which is 0.3–0.4 THz, for 1 cm wool, the accuracy is higher than 91% for either normal or 45° incidence. For 2 cm wool, the accuracy is higher than 83%. These observations validate the efficacy of the chosen frequency band in enhancing CNN's performance in seed identification tasks using terahertz imaging.

IV. CONCLUSION

In conclusion, terahertz imaging provides a viable option to help preventing GSI, while human eye observation from RGB images can be challenging when the seeds are concealed by thick layer of wool. Human eye observation is also inefficient for batch image processing. Therefore, we utilized machine learning assisted terahertz imaging to improve the accuracy of diagnosing GSI. Under normal incidence at 0.2–0.5 THz, the detecting accuracies of $> 95\%$ and $> 67\%$ are achieved when the samples are covered by 1 cm and 2 cm wool, respectively. Whereas under 45° incidence the detecting accuracy is $> 71\%$ when the sample is covered by 1 cm wool. We attribute that to scattering and attenuation from the wool, and complicated system alignment. Particularly, 0.3–0.4 THz band is demonstrated to have better performance in classification accuracy for identification of seeds. In terms of the type of seed, coriander seeds have higher validation accuracy than cumin seeds. Moreover, slightly tilted samples will not reduce the accuracy. This work provides a feasible method of batch processing of terahertz images with high

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3530900

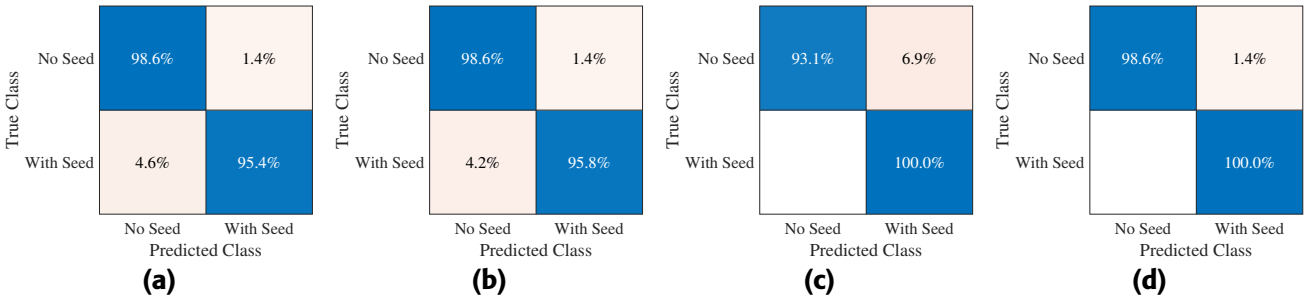


FIGURE 8. **Confusion matrix for Validation accuracy at (a) 0.2–0.5 THz, (b) 0.2–0.3 THz, (c) 0.3–0.4 THz and (d) 0.4–0.5 THz for samples in Fig. 2 (a) and (c) with 1 cm wool under normal incidence.**

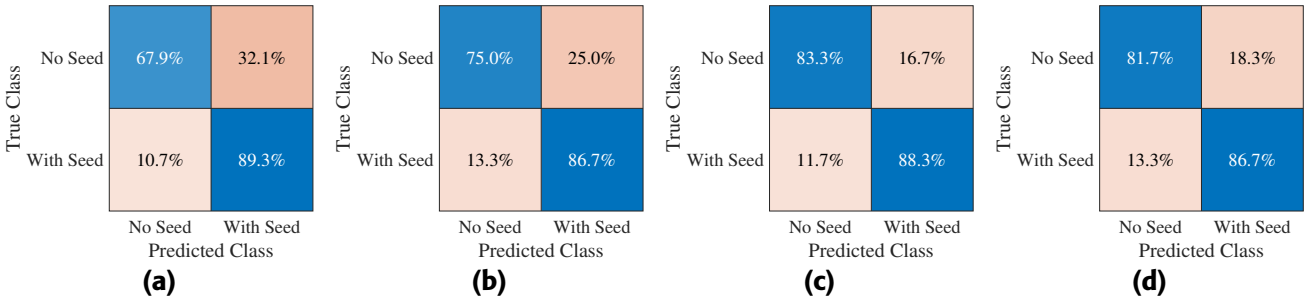


FIGURE 9. **Confusion matrix for validation accuracy at (a) 0.2–0.5 THz, (b) 0.2–0.3 THz, (c) 0.3–0.4 THz and (d) 0.4–0.5 THz for samples in Fig. 2 (a) and (c) with 2 cm wool under normal incidence.**

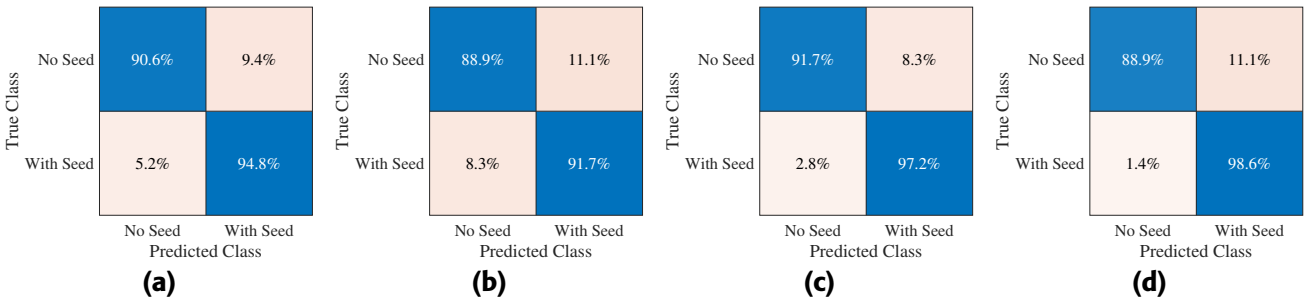


FIGURE 10. **Confusion matrix for validation accuracy at (a) 0.2–0.5 THz, (b) 0.2–0.3 THz, (c) 0.3–0.4 THz and (d) 0.4–0.5 THz for samples in Fig. 2 (a) and (f) with 1 cm wool under 45° incidence.**

accuracy and efficiency, which can be integrated into the on-field GSI detection devices.

ACKNOWLEDGMENT

The authors thank Prof. Mona Jarrahi (Department of Electrical & Computer Engineering, University of California Los Angeles) for insightful discussions. ACYG acknowledges the use of Grammarly and ChatGPT for spell check and grammar enhancement in Section III.

REFERENCES

[1] G. Lodge and B. Hamilton, "Grass seed contamination of the wool and carcases of sheep grazing natural pasture on the north-western slopes of New South Wales," *Australian Journal of Experimental Agriculture*, vol. 21, no. 111, pp. 382–386, 1981.

[2] P. Cornish and J. Beale, "Vegetable fault and grass seed infestation of sheep in New South Wales." *Journal of the Australian Institute of Agricultural Science*, vol. 40, no. 4, pp. 261–267, 1974.

[3] A. Collins, P. Fitzsummons, K. Behrendt, M. Camac, L. McCormick, R. Latta, C. John, M. Sheehan, T. Prance, A. Humphries *et al.*, "Winning against seeds: Management tools for your sheep enterprise," *Meat and Livestock Australia Ltd*, 2013.

[4] J. E. Kelly, *Seed Contamination in Sheep Carcasses by Barley Grass: An Analysis of Prevalence, Management and Economic Impact*. Charles Sturt University Wagga Wagga Australia, 2020.

[5] D. Vansteenkiste, K. Lee, and C. Lamb, "Computed tomographic findings in 44 dogs and 10 cats with grass seed foreign bodies," *Journal of Small Animal Practice*, vol. 55, no. 11, pp. 579–584, 2014.

[6] O. Cherbinsky, J. Westropp, S. Tinga, B. Jones, and R. Pollard, "Ultrasonographic features of grass awns in the urinary bladder," *Veterinary Radiology & Ultrasound*, vol. 51, no. 4, pp. 462–465, 2010.

[7] J. Cilulko, P. Janiszewski, M. Bogdaszewski, and E. Szczygielska, "Infrared thermal imaging in studies of wild animals," *European Journal of Wildlife Research*, vol. 59, pp. 17–23, 2013.

[8] T. Nagatsuma, "Terahertz technologies: present and future," *IEICE Electronics Express*, vol. 8, no. 14, pp. 1127–1142, 2011.

[9] C. Yu, S. Fan, Y. Sun, and E. Pickwell-MacPherson, "The potential of terahertz imaging for cancer diagnosis: A review of investigations to date," *Quantitative Imaging in Medicine and Surgery*, vol. 2, no. 1, p. 33, 2012.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3530900

[10] S. Fan, B. S. Ung, E. P. Parrott, V. P. Wallace, and E. Pickwell-MacPherson, "In vivo terahertz reflection imaging of human scars during and after the healing process," *Journal of Biophotonics*, vol. 10, no. 9, pp. 1143–1151, 2017.

[11] S. Thigale, Q. Wang, D. Mishra, E. M. Goldys, and S. Atakaramians, "Terahertz imaging: a diagnostic technology for prevention of grass seed infestation," *Optics Express*, vol. 31, no. 22, pp. 37 030–37 039, 2023.

[12] M. Shastri, P. Macharla, S. Thigale, E. Goldys, and S. Atakaramians, "Towards identifying grass seed infestation in the australian sheep using terahertz radiation," in *2020 Conference on Lasers and Electro-Optics Pacific Rim (CLEO-PR)*, 2020, DOI:10.1364/CLEOPR.2020.P3_17.

[13] X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, "All-optical machine learning using diffractive deep neural networks," *Science*, vol. 361, no. 6406, pp. 1004–1008, 2018.

[14] Y. Jiang, G. Li, H. Ge, F. Wang, L. Li, X. Chen, M. Lu, and Y. Zhang, "Machine learning and application in terahertz technology: A review on achievements and future challenges," *IEEE Access*, vol. 10, pp. 53 761–53 776, 2022.

[15] X. Wu, H. Liu, F. Bai, X. Lu, Y. Gao, and L. Li, "A 3D reconstruction of terahertz images based on the FCTMVSNet algorithm," *IEEE Access*, vol. 12, pp. 108 975–108 985, 2024.

[16] Y. Jiang, Y. Li, Z. Wang, F. Qi, F. Wang, Y. Liu, and T. Chen, "YOLOv8-SAB: Terahertz image detection network based on shuffle attention and YOLOv8," *IEEE Access*, vol. 12, pp. 133 328–133 338, 2024.

[17] B. Qin, Z. Li, T. Chen, and Y. Chen, "Identification of genetically modified cotton seeds by terahertz spectroscopy with MPGA-SVM," *Optik*, vol. 142, pp. 576–582, 2017.

[18] W. Liu, C. Liu, X. Hu, J. Yang, and L. Zheng, "Application of terahertz spectroscopy imaging for discrimination of transgenic rice seeds with chemometrics," *Food Chemistry*, vol. 210, pp. 415–421, 2016.

[19] J. Zhang, Y. Yang, X. Feng, H. Xu, J. Chen, and Y. He, "Identification of bacterial blight resistant rice seeds using terahertz imaging and hyperspectral imaging combined with convolutional neural network," *Frontiers in Plant Science*, vol. 11, no. 1253, 2020.

[20] Y. Jiang, H. Ge, F. Lian, Y. Zhang, and S. Xia, "Discrimination of moldy wheat using terahertz imaging combined with multivariate classification," *RSC Advances*, vol. 5, no. 114, pp. 93 979–93 986, 2015.

[21] K. Li, X. Chen, R. Zhang, and E. Pickwell-MacPherson, "Classification for glucose and lactose terahertz spectrums based on SVM and DNN methods," *IEEE Transactions on Terahertz Science and Technology*, vol. 10, no. 6, pp. 617–623, 2020.

[22] C. Wang, F. Shi, M. Zhao, J. Ao, C. Jia, and S. Chen, "Convolutional neural network-based terahertz spectral classification of liquid contraband for security inspection," *IEEE Sensors Journal*, vol. 21, no. 17, pp. 18 955–18 963, 2021.

[23] Q. Wang, A. C. Y. Goay, D. Mishra, and S. Atakaramians, "Convolutional neural network based terahertz imaging for detecting grass seed infestation," in *2024 49th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*, 2024, DOI: 10.1109/IRMMW-THz60956.2024.10697594.

[24] C. Wang, J. Qin, W. Xu, M. Chen, L. Xie, and Y. Ying, "Terahertz imaging applications in agriculture and food engineering: A review," *Transactions of the ASABE*, vol. 61, no. 2, pp. 411–424, 2018.

[25] Q. Wang, S. Afshar, H. Ebendorff-Heidepriem, and S. Atakaramians, "Realization of a single-layer terahertz magnetic mirror," *IEEE Access*, vol. 8, pp. 229 108–229 116, 2020.

[26] J. Carter, H. Lees, Q. Wang, S. J. Chen, S. Atakaramians, and W. Withayachumnakul, "Terahertz properties of common microwave dielectric materials," *Journal of Infrared, Millimeter, and Terahertz Waves*, vol. 44, no. 11, pp. 873–884, 2023.

[27] Q. Wang, Y. Deng, D. Mishra, Y. Xie, E. Aboutanios, and S. Atakaramians, "Demonstration of photonics-based D-band integrated localization and communication," *Applied Optics*, vol. 63, no. 15, pp. 4068–4076, 2024.

[28] R. Grigorev, A. Kuzikova, P. Demchenko, A. Senyuk, A. Svechkova, A. Khamid, A. Zakharenko, and M. Khodzitskiy, "Investigation of fresh gastric normal and cancer tissues using terahertz time-domain spectroscopy," *Materials*, vol. 13, no. 1, p. 85, 2019.

[29] Y. Jiang, H. Ge, F. Lian, Y. Zhang, and S. Xia, "Early detection of germinated wheat grains using terahertz image and chemometrics," *Scientific Reports*, vol. 6, no. 1, 2016.

[30] B. Li, Z.-x. Sun, A.-k. Yang, and Y.-d. Liu, "Study on detection of the internal quality of pumpkin seeds based on terahertz imaging technology," *Journal of Food Measurement and Characterization*, vol. 17, no. 2, pp. 1576–1585, 2023.

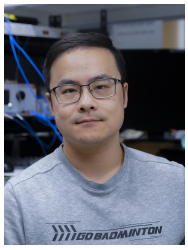
[31] J. Hu, H. Lv, P. Qiao, H. Shi, Y. He, and Y. Liu, "Research on rice seed fullness detection method based on terahertz imaging technology and feature extraction method," *Journal of Infrared, Millimeter, and Terahertz Waves*, vol. 44, no. 5, pp. 407–429, 2023.

[32] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai *et al.*, "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, 2018.

[33] S. Y. Chaganti, I. Nanda, K. R. Pandi, T. G. Prudhvith, and N. Kumar, "Image Classification using SVM and CNN," in *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA)*, 2020, DOI: 10.1109/ICCSEA49143.2020.9132851.

[34] S. K. Baranwal, K. Jaiswal, K. Vaibhav, A. Kumar, and R. Srikantaswamy, "Performance analysis of brain tumour image classification using CNN and SVM," in *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*. IEEE, 2020, pp. 537–542.

[35] R. Ribani and M. Marengoni, "A survey of transfer learning for convolutional neural networks," in *Proc. SIBGRAPI-T*. IEEE, 2019, pp. 47–57.



QIGEJIAN WANG received his B.S. in Optoelectronic Information Science and Engineering from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2017. He received his M.Phil in Electrical Engineering in 2021 from Terahertz Innovation Group, School of Electrical Engineering and Telecommunications, University of New South Wales (UNSW), Sydney, Australia. He is currently a Ph.D. candidate in the same group. He became a Graduate Student Member of IEEE in 2020.

From 2016 to 2017, Qigejian was a thesis student in the Wuhan National Laboratory for Optoelectronics, Wuhan, China. He was a research officer from 2023 to 2024, working on terahertz imaging. He also worked as lab demonstrator for the Antenna and Propagation course in UNSW. He was a recipient of UNSW Digital Grid Futures Institute 2024 Seed Funding. His current research interests include terahertz photonics, imaging and wireless communication. He is also a student member of OSA and SPIE.

AMUS CHEE YUEN GOAY received his B.Eng. (Hons) degree in Mechanical Engineering from the University of New South Wales (UNSW), Australia, in 2021. He is currently pursuing a Ph.D. in Electrical Engineering at the School of Electrical Engineering and Telecommunications, UNSW. His research interests include backscatter communication, RFID technology, energy harvesting, and system optimization.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3530900

DEEPAK MISHRA (Senior Member, IEEE) received Ph.D. in Electrical Engineering from the Indian Institute of Technology (IIT) Delhi, India in 2017. He is a Senior Lecturer at the School of Electrical Engineering and Telecommunications, University of New South Wales (UNSW) Sydney, Australia, where he joined as a Senior Research Associate in August 2019. Before that, he was a Post-Doctoral Researcher at Linköping University, Sweden, from August 2017 to July 2019. He has also been a visiting researcher at Northeastern University, USA, the University of Rochester, USA, Huawei Technologies, France, Southwest Jiaotong University, China, and Queen's University of Belfast, UK. His current research interests include energy harvesting cooperative communication networks, MIMO, backscattering, physical layer security, and signal processing and energy optimization schemes for the uninterrupted operation of wireless networks. He was a recipient of the IBM PhD Fellowship Award in 2016, the Raman Charpak Fellowship Award in 2017, the Endeavour Research Fellowship Award in 2018, and the Australian Research Council Discovery Early Career Researcher Award in 2022. Currently, he is an Associate Editor for IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING, IEEE TRANSACTIONS ON INTELLIGENT VEHICLES, IEEE WIRELESS COMMUNICATIONS LETTERS, and IEEE ACCESS.

PROFESSOR EWA M. GOLDYS is Deputy Director of the Australian Research Council Centre of Excellence in Nanoscale Biophotonics (cnbp.org.au) and Professor at the Graduate School of Biomedical Engineering, the University of New South Wales, Sydney, Australia. She is Fellow of SPIE, OSA, the Australian Academy of Technological Science and Engineering (ATSE), and winner of the 2016 Australian Museum Eureka Prize for 'Innovative Use of Technology'. She has ongoing involvement with SPIE BIOS, the world's largest international biomedical optics meeting and part of SPIE's Photonics West where she serves as one of six Track Chairs.

Her research spans the area of biomedical science, bioimaging, biosensing and materials science. She developed novel approaches to biochemical and medical sensing and deployable medical diagnostics. Current projects focus on non-invasive high content imaging of colours and patterns in cells and tissues.

SHAGHIK ATAKARAMIANS (Senior Member, IEEE) received the Ph.D. degree in electrical & electronic engineering from The University of Adelaide, Australia, in 2011, with a Certificate of Merit from the Dean. Her Ph.D. thesis received the Gertrude Rohan Memorial Prize and the 2011 University Doctoral Research Medal for outstanding research. She was with the Institute of Photonics and Optical Science (IPOS), from 2011 to 2017, and also with the Centre for Ultrahigh bandwidth Devices for Optical Systems (CUDOS), from 2012 to 2014, with the University of Sydney first as a Postdoctoral Fellow and then as a Research Fellow. She is currently an Associate Professor and founding leader of the Terahertz Innovation Group, at School of Electrical Engineering and Telecommunications, UNSW, Sydney. Her research interests include terahertz waveguides, metawaveguides, and devices. She is a Senior OPTICA (former OSA) Member. She was awarded 2024 Zhenyi Wang Award for Excellence of her outstanding contributions to the Science of Infrared, Millimeter, and Terahertz Waves.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/