



Machine learning enabled identification and real-time prediction of living plants' stress using terahertz waves

Adnan Zahid ^{a, b}, Kia Dashtipour ^b, Hasan T. Abbas ^b, Ismail Ben Mabrouk ^c, Muath Al-Hasan ^c, Aifeng Ren ^d, Muhammad A. Imran ^b, Akram Alomainy ^e, Qammer H. Abbasi ^{b, *}

^a School of Engineering and Physical Science, Heriot-Watt University, Edinburgh, EH144AS, UK

^b James Watt School of Engineering, University of Glasgow, Glasgow, G128QQ, UK

^c College of Engineering, Al-Ain University, Abu Dhabi, United Arab Emirates

^d School of Electronic Engineering, Xidian University, Xi'an, Shaanxi, China

^e School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK

ARTICLE INFO

Article history:

Received 6 December 2021

Received in revised form

24 December 2021

Accepted 4 January 2022

Available online 7 January 2022

Keywords:

Terahertz sensing

Plants health

Machine learning

ABSTRACT

Considering the ongoing climate transformations, the appropriate and reliable phenotyping information of plant leaves is quite significant for early detection of disease, yield improvement. In real-life digital agricultural environment, the real-time prediction and identification of living plants leaves has immensely grown in recent years. Hence, cost-effective and automated and timely detection of plants species is vital for sustainable agriculture. This paper presents a novel, non-invasive method aiming to establish a feasible, and viable technique for the precise identification and observation of altering behaviour of plants species at cellular level for four consecutive days by integrating machine learning (ML) and THz with a swissto12 materials characterization kit (MCK) in the frequency range of 0.75 to 1.1 THz. For this purpose, measurements observations data of seven various living plants leaves were determined and incorporate three different ML algorithms such as random forest (RF), support vector machine, (SVM), and K-nearest neighbour (KNN). The results demonstrated that RF exhibited higher accuracy of 98.87% followed by KNN and SVM with an accuracy of 94.64% and 89.67%, respectively, for precise detection of different leaves by observing their morphological features. In addition, RF outperformed other classifiers for determination of water-stressed leaves and having an accuracy of 99.42%. It is envisioned that proposed study can be proven beneficial and vital in digital agriculture technology for the timely detection of plants species to significantly help in mitigate yield and economic losses and improve crops quality.

© 2022 China Ordnance Society. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

In recent times, unceasing expansion of population has posed enormous challenges and received strong impetus to researchers and scientists in the modern agriculture discipline [1]. This mounting pressure and startling environment demands not only real-time, autonomous and non-destructive technique in perpetual plants' health monitoring at a cellular level, but also an accurate

and precise assessment of climate changes on plants leaves [2]. This significant issue driven numerous researchers and gained considerable attention of plant physiologists at all levels to establish proactive and feasible techniques for the identification of complex architecture of leaves to circumvent any unforeseen circumstances that could result in a substantial drop in crops production and economic loss in the agriculture sector. In this regard, the initial trend followed by researchers to have profoundly focused on observing the complex biological traits and microscopic behaviour of various plants leaves [3].

As envisaged, this laboratory-based technique were recognized as an ineffective, time-consuming and require manual intervention for testing and processing large number of samples [3]. Upon the

* Corresponding author.

E-mail address: qammer.abbasi@glasgow.ac.uk (Q.H. Abbasi).

Peer review under responsibility of China Ordnance Society

initial setbacks, researchers and scientists from the plant science discipline were highly motivated to suggest viable methods [4–9] to observe the morphological features of leaves at cellular level for distinct identification purpose. These techniques include fluorescence, multi-spectral or hyper-spectral imaging, magnetic resonance imaging (MRI), nuclear magnetic resonance (NMR), visible/multiband spectroscopy, and near-infrared spectroscopy (NIRS) [10–12]. Although these advanced techniques revealed distinct leaf characteristics and nutrient deficiencies in various leaves, but each technique appeared to have some limitations. For example, MRI offered high resolution but transpired as inefficient and inadequate for long-term plant physiology research due to disparaging nature. Relying on the infrared, they lacked the sensitivity to discover the spatial variability of leaves [11,12].

Hence, the researchers interest in the plant's science discipline were directed to use terahertz time-domain (THz-TDS) spectroscopy technology due to its high sensitivity, unique spectral characteristics and non-destructive nature. Initially, compared to other non-destructive technique, this approach proven to be more effective and reliable, yet again, this too were marred by some limitations and considered to be expensive, and required complex configuration of setup [12,13]. Conversely, this technology offered new opportunities in many diverse disciplines and achieved significant contributions due to its non-ionizing and less preservation radiation properties [13] as shown in Fig. 1. Several applications where THz achieved significant contributions such as security imaging of unseen hazard items, medical imaging for non-intrusive dental and skin-care treatment, quality control of food, contactless imaging for conservation of sculptures and manuscripts, material characterization and telecommunications [13–15]. Despite its extensive utility and notable contributions to the aforementioned diverse applications, researchers believed its potential in modern agriculture technology has yet to be fully explored [13].

These undermine facts and prevailing challenges have captivated researchers from different disciplines and markedly stipulate an automated real-time, proactive and pragmatic technique approach for different plants species identification [16–18]. Unlike fruits and flowers, leaves are accessible all year around. As a result, effective plant species identification is critical for ensuring sustainable cultivation and large-scale crops production by closely monitoring the species populations and assessing the impact of

climate change on species distribution. Moreover, it is seemed to be highly beneficial to accelerate the recognition process of distinct feature of leaves, making it consumable by non-specialists, especially when considering the incessant damage of plant biodiversity [16–18].

This paper presents an innovative, real-time and non-invasive technique for instant identification, lifespan of plant species at cellular level using THz waves enabled by machine learning (ML) approach in an automated fashion. Up till now machine learning applications have attained substantial contributions in numerous scientific domains, including health management and safety, food utilization, security, weather forecasting, pharmaceutical drugs, and financial sector [16,19]. Evidence from multi-disciplinary disciplines and researchers suggest that it has the potential to revolutionize the modern agriculture system by raising crops productivity and optimize the financial gains in agriculture sector. Furthermore, the amalgamation of ML with THz technology can unravel and discover innovative possibilities for various species identification, as well as comprehend the intensive data processing in agricultural environment [19].

In this study, the underlying aim is to develop a simple and rapid framework to recognize the correct species in conjunction with lifespan of plant leaves using THz waves.

For this purpose, we obtained transmission response S_{21} of seven living plants leaves for four days, applied feature extraction and selection technique to remove the undesired data, and finally performed the supervised ML classification for species identification and plant's stress detection at cellular level. It is strongly envisioned that findings of this study will benefit growers, cultivators and consider it as versatile, extensible, and provide accurate phenotyping information to researchers and plant biologists. The remainder of the paper is organized as follows: "Methodology" includes the description of setup and sample details. Section 2 outlines the compilation and pre-processing of data. Section 2 describes feature extraction procedure. Section 3 demonstrates the classification technique and optimal feature selection process. Finally section 4 presents the results analysis and discussions, followed by conclusion in section 5.

2. Methodology

2.1. System setup

In this study, scattering parameters of seven living plants were determined by using Material Characterization Kit (MCK) swissto12 that operated in the frequency region of 0.75–1.1 THz [20]. The MCK was attached to a Keysight Technologies N5224A microwave network analyzer (NA). Prior to measurements, short-open-load-thru (SOLT) calibration technique, which is two-port, was used to lessen any undesired noise in system setup. Considering the structural integration and morphological characteristics of leaves, Polytetrafluoro-ethylene (PTFE) caps were mounted internally to the waveguide that could provide appropriate compression and stability to the specimens as shown in Fig. 2.

2.2. Preparation of samples

This work includes the measurements of seven various kinds of plants namely as basil, coriander, parsley, baby-leaf, coffee, pea-shoot, and babyspinach. All these various plants leaves were fully grown and nurtured in Rouken Glen Farm, East Renfrewshire, Glasgow. As per the status of these plants leaves, aforementioned leaves were properly cultivated and nurtured, and showed no signs of damage after proper examination of farmers. They were properly placed under the environment temperature of 18 ± 0.1 °C, in order

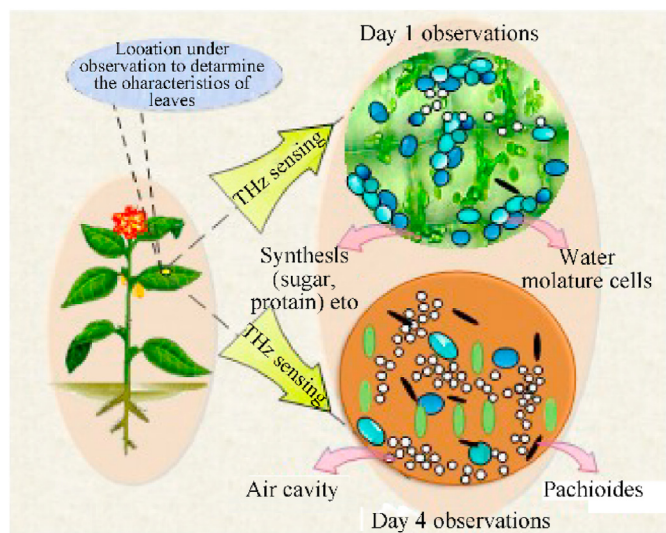


Fig. 1. Observing the morphological characteristics of leaves using THz sensing at cellular level on different days.

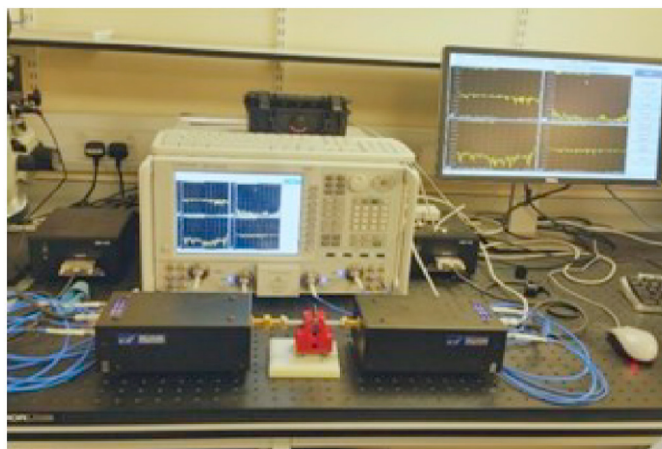


Fig. 2. Snapshot of setup showing a Swissto12 System operates in a frequency range from 0.75 to 1.1 THz.

to maintain their healthy status for measurements. Moreover, in this study, the weight and thickness were continually observed for four consecutive days after every 2 h using precision electronic scale and vernier calliper. Over the span of four days, spectral variations appeared that demonstrated the reduction of weight and thickness in them. These eaves were closely examined at three different locations and at every location, four distinct orientations were taken to observe any irregularity in the epidermis of leaves that could show any disparity in the scattering response of leaves.

2.3. Data collection and pre-processing procedure

In this study, seven fresh living plants leaves, namely as baby-leaf, basil, coriander, coffee-arabica, parsley, pea-shoot and baby-spinach were considered for performing measurements. All measurements were conducted in Terahertz Frequency Laboratory (TFL) at the University of Glasgow for four consecutive days. In this study, the key idea and focus was to observe the transmission coefficients (S_{12}, S_{21}) and reflection (S_{11}, S_{22}) by placing all distinct leaves between the two wave-guides as shown in Fig. 2. In this study, only transmission response of leaves was considered for the real-time identification and detection of various living plant leaves' stress as shown in Fig. 2. This process continued for four days and in this period, all specimens were examined after a span of every 3 h to ensure that maximum information could be obtained in order to closely and warily observe any transformations of leaves at cellular level. A key challenge whilst acquiring the data collection of distinct leaves, was to determine the same location for examining the internal morphological composition of leaves. As a result, special care was taken to ensure that the significant information and diminutive variations in complex biological traits of leaves could be captured.

From Fig. 3, it is depicted that leaves showed distinctive characteristics depending on the existence of water content WC in tissue on day 1 and as days passed by, as anticipated, distinguished response of leaves is obtained reflecting the variations occurred at cellular level detected by THz waves. distinguished from unusual response of leaves. These observations reflect the presence of WC perceived over the course of four days measurements. Table 1 illustrates the observations of different leaves obtained using MCK, reflecting the existence of WC in leaves.

By observing the leaves' transmission response over the course of four days, it is evident each leaf has shown distinctive characteristics and behaviour that appeared to be occurred at cellular level and is markedly visible in THz region. From Fig. 3, it is also

believed that all leaves exhibited a decay in WC that eventually enhanced the transmission response leaves. In addition, detection of distinct leaves is relatively more challenging in the range of 0.75–0.85 THz due to overlapping of response. Hence, it is established that distinguished response of leaves was obtained as the frequency increases, reflecting the composition and variance presence of nutrients in the leaves.

2.4. Feature extraction technique

For the feature extraction procedure, three domain features including time-frequency, time and frequency domain features were considered [21]. While taking the measurements of leaves using Swissto12 (MCK), it was noticed that observations collected in the region of 0.75–0.83 THz exhibited superfluous behaviour. The outcome of this unwanted and erratic behaviour could have produced spurious classification results for different classifiers, resulting in forge detection and identification of correct species of plant leaves. Moreover, this could lead inaccurate information to growers and cultivators to determine which species demands how much of nutrients and WC in leaves to maintain their healthy status and sustain crops productivity. As a result, it was significant to determine the sensitive response region (SRR) so that maximum meaningful observations could be obtained for the classification model as shown in Fig. 5. Since the MCK operated in the frequency domain, therefore observations obtained from MCK were converted into time and time-frequency domain to obtain features extraction technique that could be useful for classifiers performance. In the next section, both domain features are discussed in detail.

2.5. Frequency domain features

Since the observations were collected in frequency domain using MCK therefore, these observations could be used to extract the relevant features from the SRR region as shown in Fig. 4. SRR assisted in the accurate identification of plant species as well as water-stressed leaves because various leaves exhibited distinct characteristics, which were examined for feature extraction process. Thus, for all four days, five window bins with a width of 20 were considered for feature extraction, beginning from the frequency range of 0.80–1.07 THz. To reach this objective, both Power and Cross spectral densities were studied and is given as 1 and 2 [15,22]:

$$\text{Var}(Z_{mm}(b) = \frac{1}{W} E(Z_v^u(b)^* \cdot Z_v^u(b))) \tag{1}$$

$$\text{max}(Z_{mm}(b) = \text{max} \frac{1}{W} EK(b)^* \cdot Z_v^u(b)) \tag{2}$$

2.6. Time domain features

The primary objective of studying statistical characteristics was to investigate the spectral features of time series of THz pulses related to the existence of WC and modifications in morphological features of leaves at cellular level over the course of four days. In order to acquire statistical features, it was therefore, essential to translate observations obtained in frequency into time domain to profoundly study the significant and meaningful THz pulse of distinct leaves. The statistical features including standard deviation, skewness, kurtosis median, absolute deviation, Interquartile Range (IQR), 75th percentile (Q3), 25th percentile (Q1) and Pearson correlation coefficient (PCC) [23–25]. Wherein, mean and standard

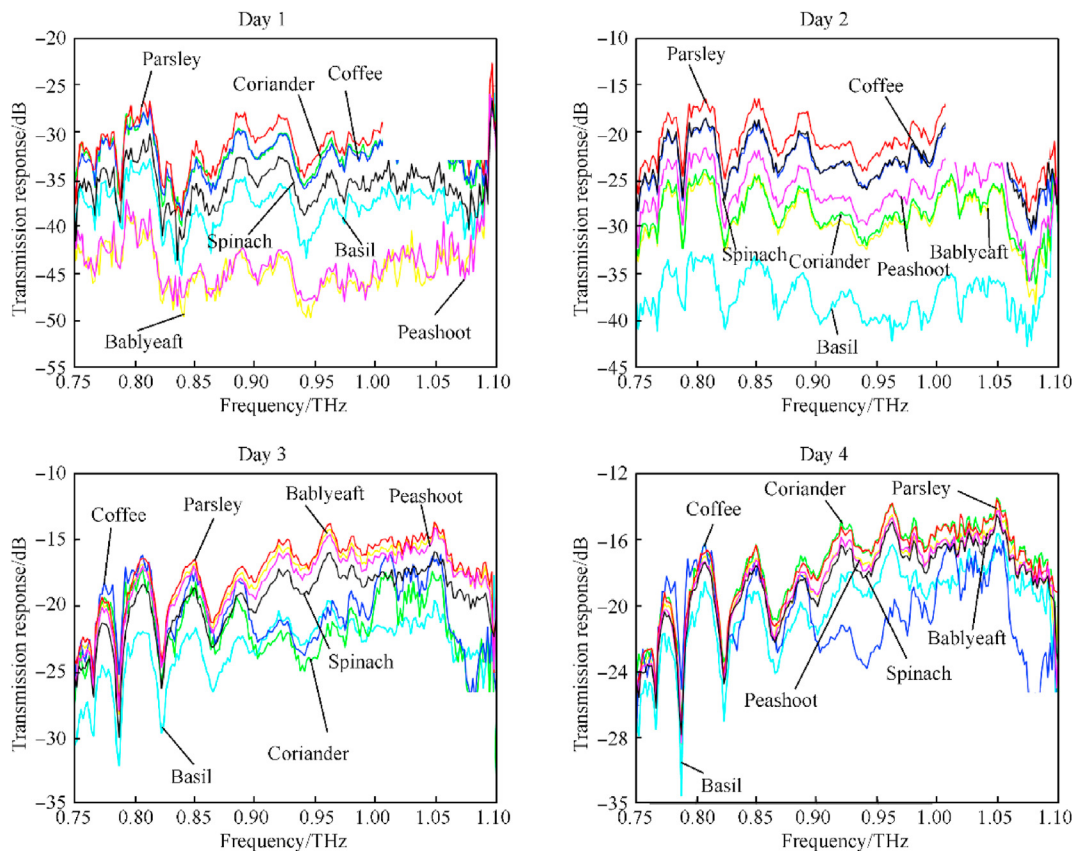


Fig. 3. Transmission response of seven living plants leaves for all different days.

Table 1
Number of Observations of different leaves for four successive days.

Leaves	No. of observations
Baby-leaf	108
Basil	108
Coriander	108
Coffea arabica	108
Parsley	108
Pea-shoot	108
Baby Spinach	108

deviation were notably suitable for providing valuable details on distribution of the data. Skewness yielded relevant and useful data about the inconsistencies of the inspected region and its dissemination around its mean scale region [23–25]. Furthermore, kurtosis indicated a uniform distribution of data. Q1 and Q3 described the dispersion of observed data on both sides of the median. PCC was used to establish the linear correlation between the reference signal and time-domain waveforms of sample [16]. All these statistical features played a vital role in identifying the correct and relevant features which ultimately reduced the computational time for overall classification process.

3. Classification and selection of optimal parameters

In order to achieve the classification accuracy, three classifiers namely Random Forest (RF), support vector machine (SVM), and K-nearest neighbour (KNN) were employed to determine characteristics behaviour of leaves for four successive days. In addition, the

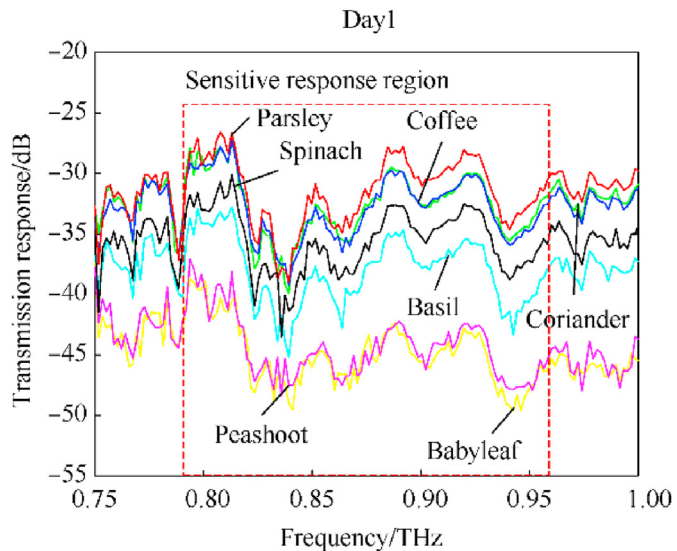


Fig. 4. Recognition of sensitive region to consider only desired and meaningful features for the feature extraction.

performance of three aforementioned classifiers were also tested for the precise identification of leaves by observing their physiological and biological characteristics over the span of four consecutive days. The performance of three classifiers were tested and analysed by selecting appropriate parameters to yield optimum results in relation to accurate recognition of leaves and as well as water stressed leaves. In this regard, for SVM, two parameters were

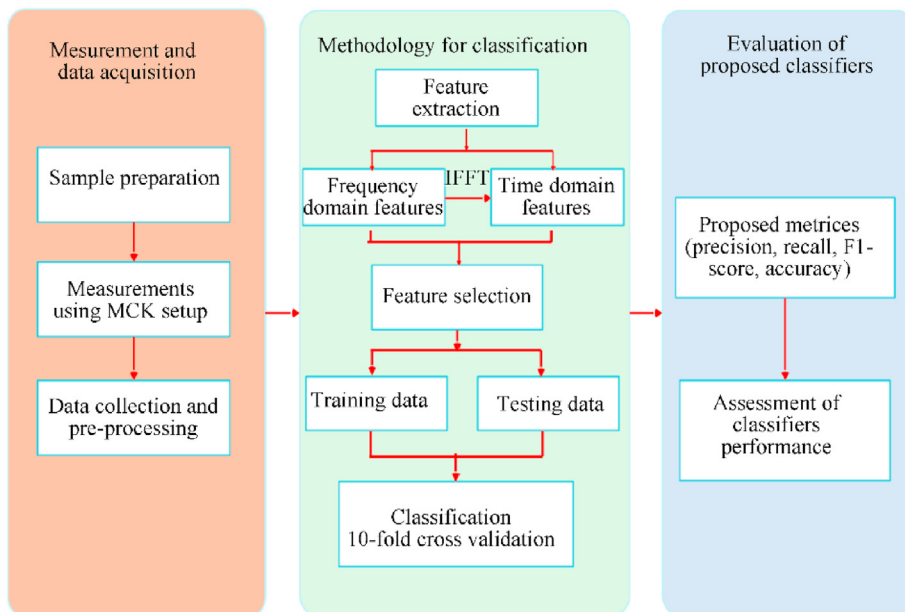


Fig. 5. The methodological approach of proposed algorithm implementation process.

examined, namely as the optimum parameters of cost (C) and kernel width parameter γ to obtain the optimized SVM algorithm. For this purpose, random values were gauged and ultimately '1.5' considered to be appropriate value for 'C' and "0.36" was selected for γ [16,26,27].

The tuning of a k -sample parameter played an important role in achieving the ultimate efficiency of KNN classifier. Hence, set of values was tested and finally, nearest number of k and distance metric was set to 5 after analysing the range of k (1:10) [26,27]. Furthermore, observations data was divided into 70% and 30% training and testing data, respectively. All three suggested classifiers were trained by using a 10-fold cross validation to achieve the validation accuracy for the precise recognition of species by analysing their physiological and morphological characteristics. Owing to the loss of WC and other nutrients in leaves, variations evidently appeared in the physiological and sensory characteristics in four days assessments.

3.1. Feature selection

In this study, the core idea was to eliminate any superfluous or undesired features with the usage of feature selection technique. This action would ultimately reduce the computational load and result in substantial improvement in classification accuracy in shorter time frame. For this purpose, three feature selection algorithms such as sequential backward selection (SBS), sequential forward selection (SFS), and Relief based selection algorithm (Relief-F) were considered to be most suitable since they have been widely used to execute the feature selection process [28–30]. SFS is one such method, which begins with an empty set of features and append them by incorporating the relevant and noticeable features, leading to an improvement in overall accuracy. On the other hand, SBS functions in reverse order as it commences with totally filled features and get rid of unparalleled features in every level by introducing standard condition until the pre-specified features are permitted [28–30].

3.2. Classifier performance evaluation using metrics

In this work, the purpose of study was to examine the efficiency of classifiers and also to detect any potential mis-classification that could have produced inaccurate information about the detection of leaves. In addition, it was also aimed to determine whether the plants' leaves were supplied with adequate amount of WC and nutrients to ensure and maintain their physiological growth. Furthermore, in this section, the performance of all proposed classifiers was also evaluated by using three commonly metrics such as, precision, recall (also known as true positive), and F1-score [31] as shown in Table 2. Here, precision metrics were employed to evaluate the precision of the classifications relative to all other classifications. In addition, recall or sensitivity values showed the possibility of occurring accurate classification of categorised classes from the remaining classes. Finally, F1-score was used to obtain the average between the Precision and Recall metrics [31]. In this study, the key objective of using these commonly agreed metrics was to detect any potential mis-classification, resulting in inaccurate details about the presence of WC and nutrients in leaves.

4. Results and discussions

4.1. Classifiers performance for accurate detection of leaves

The performance of three proposed algorithms including RF, KNN and SVM was assessed in this section by employing commonly quality metrics for the automated identification of different leaves possessing distinctive characteristics and features with an amalgamation of ML and THz. The purpose of employing commonly accepted quality metrics was also to detect any mis-classification that could take place, resulting in erroneous information about characteristics and identification of leaves. From Tables 2–4, it can be noticed RF substantially outperformed other classifiers in terms of accurate detection of all seven leaves by observing their internal morphological characteristics using THz for four days, revealing the fresh moistness and moldiness of leaf.

Furthermore, these outcomes also revealed that for SVM, it was

Table 2
Classification accuracy of RF for the accurate recognition of leaves by applying Ten-fold cross validation.

Classification performance by applying tenfold cross validation				
Sample	Quality metrics			Accuracy/%
	Precision	Recall	F1-score	
Coffee	0.97	0.96	0.97	98.87
Pea-shoot	0.97	0.96	0.96	
Spinach	0.97	0.96	0.96	
Coriander	0.97	0.96	0.95	
Basil-leaf	0.97	0.96	0.96	
Baby-leaf	0.97	0.96	0.96	
Parsley	0.97	0.96	0.97	

Table 3
Classification accuracy of SVM for the accurate recognition of leaves by applying Ten-fold cross validation.

Classification performance by applying tenfold cross validation				
Sample	Quality metrics			Accuracy/%
	Precision	Recall	F1-score	
Coffee	0.89	0.98	0.93	89.67
Pea-shoot	0.91	0.78	0.84	
Spinach	0.89	0.86	0.88	
Coriander	0.9	0.98	0.94	
Basil-leaf	0.86	0.92	0.89	
Baby-leaf	0.89	0.8	0.84	
Parsley	0.93	0.96	0.95	

Table 4
Classification accuracy of KNN for the accurate recognition of leaves by applying Ten-fold cross validation.

Classification performance by applying tenfold cross validation				
Sample	Quality metrics			Accuracy/%
	Precision	Recall	F1-score	
Coffee	0.96	0.99	0.97	94.64
Pea-shoot	0.94	0.89	0.92	
Spinach	0.94	0.95	0.94	
Coriander	0.98	0.99	0.98	
Basil-leaf	0.93	0.95	0.94	
Baby-leaf	0.93	0.88	0.9	
Parsley	0.96	0.97	0.96	

slightly challenging to identify various leaves due to the similar sensory characteristics and somewhat identical appearance such as coriander, parsley. However, as observed in the results, KNN displayed considerable performance relative to SVM, given the physiological and biological characteristics of leaves. Moreover, the results also depicted that all three classifiers considered only notable and desired features which inevitably reduced the computation time for the execution of classifiers. Hence, by selecting the desired and relevant features not only reduced the estimation time for execution but also enhanced the performance of all three classifiers.

4.2. Classifiers performance for observing characteristics on different days

The purpose of this research was to minutely observe the behaviour and physiological traits of all seven living plants leaves over the course of four days. In this regard, an algorithm was proposed for three different classifiers and their performance was also assessed by quality metrics. From Table 5, it can be seen that for coffee, RF outperformed other classifiers to achieve performance of different days classification, revealing the precise estimation of WC

in coffee leaf. Both KNN and SVM displayed relatively less accuracy in precise estimation of WC in coffee as depicted in Table 5. Moreover, the assessment of all proposed classifiers model showed a noticeable performance by depicting accuracy of 90% for all classifiers from day 1 to day 4, clearly revealing the freshness and staleness of a leaf. For both spinach and peashoot leaves as shown in Tables 6 and 7, RF exhibited distinctive performance compare to KNN and SVM for the presence of WC in leaf. It was also perceived that in both KNN and SVM, variability of detecting the WC on both day 2 and 3 found in the range of 0.55–0.86 THz, resulting in low precision, while RF displayed considerable accuracy, clearly indicating a high and low amount of WC in leaves on days 1 and 4.

As shown in Table 10, the performance of all classifiers can be assessed for observing diminutive changes in cellular level in coriander leaf, which is due to the WC evaporation as compared to other leaves. This also revealed that parameters selected for all three classifiers evidently exhibited a meaningful accuracy ranging from 94.34% to 99.97%, identifying the presence of WC in leaf for different days. Upon a close analysis of both baby and basil leaf as presented in Tables 9 and 11, it can be seen that performance of all three classifiers for the detection of WC was not satisfactory. In comparison to KNN and SVM, RF showed distinguished

Table 5
Performance of all three classifiers for observing the WC variations in coffee leaf for four days.

Classification performance by applying tenfold cross validation coffee					
Classifier	Days	Quality metrics			Accuracy/%
		Precision	Recall	F1-score	
Random forest	Day 1	1	1	1	93.09
	Day 2	0.89	0.89	0.89	
	Day 3	0.86	0.86	0.86	
	Day 4	0.97	0.97	0.97	
KNN	Day 1	1	1	1	91.77
	Day 2	0.88	0.87	0.87	
	Day 3	0.83	0.84	0.84	
	Day 4	0.96	0.96	0.96	
SVM	Day 1	1	1	1	92.72
	Day 2	0.93	0.85	0.89	
	Day 3	0.83	0.89	0.86	
	Day 4	0.96	0.96	0.96	

Table 6
Performance of all three classifiers for observing the WC variations in spinach leaf for four days.

Classification performance by applying tenfold cross validation spinach					
Classifier	Days	Quality metrics			Accuracy/%
		Precision	Recall	F1-score	
Random forest	Day 1	1	1	1	93.90
	Day 2	0.91	1	0.95	
	Day 3	0.91	0.88	0.9	
	Day 4	0.93	0.88	0.91	
KNN	Day 1	1	1	1	77.96
	Day 2	0.83	0.9	0.86	
	Day 3	0.63	0.59	0.61	
	Day 4	0.64	0.63	0.64	
SVM	Day 1	1	0.99	1	73.75
	Day 2	0.78	0.84	0.81	
	Day 3	0.71	0.75	0.73	
	Day 4	0.55	0.87	0.67	

Table 7
Performance of all three classifiers for observing the WC variations in peashoot for four days.

Classification performance by applying tenfold cross validation peashoot					
Classifier	Days	Quality metrics			Accuracy/%
		Precision	Recall	F1-score	
Random forest	Day 1	1	1	1	90.97
	Day 2	0.99	0.95	0.97	
	Day 3	0.85	0.83	0.84	
	Day 4	0.81	0.87	0.84	
KNN	Day 1	0.98	1	0.99	77.20
	Day 2	0.95	0.93	0.94	
	Day 3	0.59	0.57	0.58	
	Day 4	0.78	0.6	0.58	
SVM	Day 1	1	0.98	0.99	71.91
	Day 2	0.96	0.93	0.94	
	Day 3	0.63	0.62	0.63	
	Day 4	0.62	0.66	0.63	

classification accuracy for the precise estimation of WC in both baby and basil leaves.

Considering the internal morphological configuration of both baby and basil leaves, it was observed that the performance of classifiers on both days 2 and 3 was reduced than to day 1 and day 4. On closer observation, this might be attributed to the biological variations that occurred in tissue of leaves due to the presence of water in the range of 30%–50%. In this case, the factors that have influenced the classification accuracy in terms of precisely detecting the presence of WC due to the internal physiological and

biological characteristics which have progressively changed over the course of days, resulting in precipitous evaporation of WC from leaves.

Lastly, by observing the water status in parsley leaf for four consecutive days, it was established from [Table 8](#) that RF once again demonstrated 100% performance producing 100% as compared to SVM and KNN, yielding 79.66% and 97.89%, respectively. The significant performance of RF also suggested that both sensory characteristics and biological traits that occurred in parsley leaf at cellular level were discernibly recognized due to the evaporation of

Table 8
Performance of all three classifiers for observing the WC variations in parsley leaf for four days.

Classification performance by applying tenfold cross validation parsley					
Classifier		Quality metrics			Accuracy/%
Algorithm	Days	Precision	Recall	F1-score	
Random forest	Day 1	1	1	1	100
	Day 2	1	1	1	
	Day 3	1	1	1	
	Day 4	1	1	1	
KNN	Day 1	1	0.91	0.96	97.89
	Day 2	0.93	0.94	1	
	Day 3	1	1	1	
	Day 4	1	1	1	
SVM	Day 1	1	1	1	79.66
	Day 2	0.91	0.93	0.92	
	Day 3	0.63	0.67	0.65	
	Day 4	0.64	0.79	0.72	

Table 9
Performance of all three classifiers for observing the WC variations in baby leaf for four days.

Classification performance by applying tenfold cross validation baby					
Classifier		Quality metrics			Accuracy/%
Algorithm	Days	Precision	Recall	F1-score	
Random forest	Day 1	1	1	1	79.70
	Day 2	0.99	1	0.99	
	Day 3	0.59	0.65	0.62	
	Day 4	0.6	0.54	0.57	
KNN	Day 1	1	1	1	80.18
	Day 2	1	1	1	
	Day 3	0.61	0.59	0.6	
	Day 4	0.6	0.62	0.61	
SVM	Day 1	1	0.56	0.72	76.18
	Day 2	0.99	0.91	0.95	
	Day 3	0.71	0.82	0.76	
	Day 4	0.66	0.65	0.76	

Table 10
Performance of all three classifiers for observing the WC variations in coriander leaf for four days.

Classification performance by applying tenfold cross validation coriander					
Classifier		Quality metrics			Accuracy/%
Algorithm	Days	Precision	Recall	F1-score	
Random forest	Day 1	1	1	1	99.89
	Day 2	1	1	1	
	Day 3	1	1	1	
	Day 4	1	1	1	
KNN	Day 1	1	1	1	99.97
	Day 2	1	1	1	
	Day 3	1	1	1	
	Day 4	1	1	1	
SVM	Day 1	1	1	1	94.34
	Day 2	1	0.78	0.88	
	Day 3	0.9	1	0.95	
	Day 4	0.89	1	0.94	

WC for four days, showing 100% classification accuracy. In addition, KNN also illustrated a strong capability to determine the leaf water status in drought stress for all four days. On the contrary, SVM exhibited a lower accuracy due to the variations occurred on day 3 and 4, respectively.

In the current environment, availability of fertile land is limited due to extreme climate transformations, therefore, it is strongly envisioned that the proposed classification methodology for plants specimen identification and continuous real-time observation of water-stressed leaves can be significant in the implementation of

digital agricultural system to improve crops productivity by proactive monitoring of health status of leaves. Moreover, the availability of suggested study can be seen beneficial, meaningful, and as promising candidate in providing proactive early alerts of plants drought stresses at an early stage while ensuring adequate use of water usage in the field of plant biology. This will also encourage cultivators, horticulturists, to take effective and satisfactory measures to maintain the healthy status of plants by timely monitoring their demand of water distribution and nutrients.

Table 11
Performance of all three classifiers for observing the WC variations in basil leaf for four days.

Classification performance by applying tenfold cross validation basil leaf					
Classifier	Days	Quality metrics			Accuracy/%
		Precision	Recall	F1-score	
Random Forest	Day 1	0.85	0.71	0.77	89.03
	Day 2	0.75	0.88	0.81	
	Day 3	1	0.98	0.99	
	Day 4	0.98	1	0.99	
KNN	Day 1	0.64	0.59	0.61	78.94
	Day 2	0.62	0.67	0.64	
	Day 3	1	0.9	0.95	
	Day 4	0.91	1	0.95	
SVM	Day 1	0.6	0.58	0.59	70.60
	Day 2	0.6	0.62	0.61	
	Day 3	0.77	0.89	0.83	
	Day 4	0.87	0.73	0.79	

5. Conclusions

This paper highlights the emergence of terahertz (THz) technology enabled by machine learning (ML) non-invasive approach for the precise and real-time identification of various plants specimen in an automated fashion. For this purpose, transmission response of seven distinct plant leaves were measured for four consecutive days. To perform the classification, useful and meaningful features were selected by performing feature selection technique in order to enhance the classification accuracy for the prediction of accurate species of plant leaves. The results drastically showed improvement after identifying the important features, yielded significant information about the water-stressed leaves by determining cross-validation methodology. In addition, the computation time was also improved that ultimately enhanced the execution time of all three classifiers. It was noticed that for real-time prediction and identification of different leaves, RF exhibited higher accuracy of 98.87% followed by KNN and SVM with an accuracy of 94.64% and 89.67%, respectively, by observing their morphological characteristics at cellular level. In addition, RF outperformed other classifiers with precision accuracy of 99.42% for determination of water-stressed leaves.

In an environment, where climate transformations is growing and limited availability of fertile land, the proposed study has the strong potential to provide valuable recommendations and observations to horticulturists and botanists to develop a smart, sustainable digital agricultural technology by providing appropriate phenotyping information of plant leaves in an automated fashion, which is of great significant to improve the productivity of crops.

Author contributions

Conceptualization, A. Z., K. D., H. T. A., and Q. H. A.; software, A. Z., K. D. and A. R.; resources, A. Z., Q. H. A.; writing-original draft preparation, A. Z. and K. D.; writing-review and editing, H.T.A, I.B. M., M. H., A. R., A. A., and Q. H. A.; supervision, Q. H. A. and M. A. I.; project administration, Q. H. A.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was funded under EPSRC DTA studentship which

is awarded to A. Z. for his PhD. Research Council (DTG EP/N509668/1 Eng).

References

- [1] Roberts CJ. An integrated approach to function annotation in the histidine phosphatase superfamily. 2015.
- [2] Lowe A, Harrison N, French AP. Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. *Plant Methods* 2017;13(1):80.
- [3] Fariñas MD, Jimenez-Carretero D, Sancho-Knapik D, PegueroPina JJ, Gil-Pelegri'n E, Alvarez-Arenas TG. Instantaneous and non-destructive relative water content estimation from deep learning applied to resonant ultrasonic spectra of plant leaves. *Plant Methods* 2019;15(1):1–10.
- [4] Afsharinejad A, Davy A, Naftaly M. Variability of terahertz transmission measured in live plant leaves. *Geosci Rem Sens Lett IEEE* 2017;14(5):636–8.
- [5] Born N, Behringer D, Liepelt S, Beyer S, Schwerdtfeger M, Ziegenhagen B, Koch M. Monitoring plant drought stress response using terahertz time-domain spectroscopy. *Plant Physiol* 2014;164(4):1571–7.
- [6] Federici JF, Schulkin B, Huang F, Gary D, Barat R, Oliveira F, Zimdars D. Thz imaging and sensing for security applications—explosives, weapons and drugs. *Semicond Sci Technol* 2005;20(7):S266.
- [7] Naftaly M, Miles RE. Terahertz time-domain spectroscopy for material characterization. *Proc IEEE* 2007;95(8):1658–65.
- [8] Jordens C, Scheller M, Breitenstein B, Selmar D, Koch M. Evaluation of leaf water status by means of permittivity at terahertz frequencies. *J Biol Phys* 2009;35(3):255–64.
- [9] Liew OW, Chong PCJ, Li B, Asundi AK. Signature optical cues: emerging technologies for monitoring plant health. *Sensors* 2008;8(5):3205–39.
- [10] Torres V, Palacios I, Iriarte JC, Liberal I, Santesteban LG, Miranda C, Royo JB, Gonzalo R. Monitoring water status of grapevine by means of thz waves. *J Infrared, Millim Terahertz Waves* 2016;37(5):507–13.
- [11] Santesteban LG, Palacios I, Miranda C, Iriarte JC, Royo JB, Gonzalo R. Terahertz time domain spectroscopy allows contactless monitoring of grapevine water status. *Front Plant Sci* 2015;6:404.
- [12] Song Z, Yan S, Zang Z, Fu Y, Wei D, Cui H-L, Lai P. Temporal and spatial variability of water status in plant leaves by terahertz imaging. *IEEE Trans Terahertz Sci Technol* 2018;8(5):520–7.
- [13] Zahid A, T Abbas H, Imran MA, Qaraqe KA, Alomainy A, Cumming DR, Abbasi QH. Characterization and water content estimation method of living plant leaves using terahertz waves. *Appl Sci* 2019;9(14):2781.
- [14] Zahid A, Abbas HT, Sheikh F, Kaiser T, Zoha A, Imran M, Abbasi QH. Monitoring health status and quality assessment of leaves using terahertz frequency. In: 2019 IEEE international symposium on antennas and propagation and USNC-ursi radio science meeting. IEEE; 2019. p. 379–80.
- [15] Zahid A, Abbas HT, Ren A, Zoha A, Heidari H, Shah SA, Imran MA, Alomainy A, Abbasi QH. Machine learning driven non-invasive approach of water content estimation in living plant leaves using terahertz waves. *Plant Methods* 2019;15(1):138.
- [16] Ren A, Zahid A, Zoha A, Shah SA, Imran MA, Alomainy A, Abbasi QH. Machine learning driven approach towards the quality assessment of fresh fruits using non-invasive sensing. *IEEE Sensor J* 2019;20(4):2075–83.
- [17] Huang Y, Ren Z, Li D, Liu X. Phenotypic techniques and applications in fruit trees: a review. *Plant Methods* 2020;16(1):1–22.
- [18] Rzanny M, Seeland M, Wäldchen J, Mäder P. Acquiring and preprocessing leaf images for automated plant identification: understanding the tradeoff between effort and information gain. *Plant Methods* 2017;13(1):1–11.
- [19] Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D. Machine learning in agriculture: a review. *Sensors* 2018;18(8):2674.
- [20] Swissto12 URL. <http://www.swissto12.com>.

- [21] Li H, Yuan D, Wang Y, Cui D, Cao L. Arrhythmia classification based on multi-domain feature extraction for an ecg recognition system. *Sensors* 2016;16(10):1744.
- [22] Von Storch H, Zwiers FW. *Statistical analysis in climate research*. Cambridge university press; 2001.
- [23] Chen H, Chen X, Ma S, Wu X, Yang W, Zhang W, Li X. Quantify glucose level in freshly diabetic's blood by terahertz time-domain spectroscopy. *J Infrared, Millim Terahertz Waves* 2018;39(4):399–408.
- [24] Yin X, Hadjiloucas S, Zhang Y, et al. Classification of thz pulse signals using two-dimensional cross-correlation feature extraction and non-linear classifiers. *Comput Methods Progr Biomed* 2016;127:64–82.
- [25] Dutta S, Chatterjee A, Munshi S. Correlation technique and least square support vector machine combine for frequency domain based ecg beat classification. *Med Eng Phys* 2010;32(10):1161–9.
- [26] Li C, Wang J, Wang L, Hu L, Gong P. Comparison of classification algorithms and training sample sizes in urban land classification with landsat thematic mapper imagery. *Rem Sens* 2014;6(2):964–83.
- [27] Thanh Noi P, Kappas M. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery. *Sensors* 2018;18(1):18.
- [28] Gürbüz SZ, Erol B, Çağlıyan B, Tekeli B. Operational assessment and adaptive selection of micro-Doppler features, *IET Radar. Sonar Navig* 2015;9(9): 1196–204.
- [29] Pohjalainen J, Räsänen O, Kadioglu S. Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits. *Comput Speech Lang* 2015;29(1):145–71.
- [30] Feizi-Derakhshi M-R, Ghaemi M. Classifying different feature selection algorithms based on the search strategies. In: *International conference on machine learning, electrical and mechanical engineering*; 2014. p. 17–21.
- [31] Gibson RM, Amira A, Ramzan N, Casaseca-de-la Higuera P, Pervez Z. Multiple comparator classifier framework for accelerometer-based fall detection and diagnostic. *Appl Soft Comput* 2016;39:94–103.