ECTI Transactions on Computer and Information Technology
Journal homepage: https://ph01.tci-thaijo.org/index.php/ecticit/
Published by the ECTI Association, Thailand, ISSN: 2286-9131

Non-Destructive Inspection of Tile Debonding by DWT and MFCC of Tile-Tapping Sound with Machine versus Deep Learning Models

Jantana Panyavaraporn[1] and Paramate Horkaew[2]

ABSTRACT

One of the essential processes of construction quality control is tile bonding inspection. Hollows beneath tile tessellation can lead to unbounded or completely broken tiles. An interior inspector typically used a hollow-sounding technique. However, it relies on skill and judgment that greatly vary among individuals. Moreover, equipment and interpretation are difficult to calibrate and standardize. This paper addresses these issues by employing machine-learning strategies for tile-tapping sound classification. Provided that a tapping signal was digitally acquired, the proposed method was fully computerized. Firstly, the signal was analyzed and its wavelets and MFCC were extracted. The corresponding spectral features were then classified by SVM, k-NN, Naïve Bayes, and Logistic Regression algorithm, in turn. The results were subsequently compared against those from a previous works that employed a deep learning strategy. It was revealed that when the proposed method was properly configured, it required much less computing resources than the deep learning based one, while being able to distinguish dull from hollow sounding tiles with 93.67% accuracy.

Article information:

Keywords: Tile Tapping Sound, Support Vector Machine, Logistic Regression, K-Nearest Neighbors, Naive Bayes

Article history:
Received: August 14, 2023
Revised: December 7, 2023
Accepted: January 4, 2024
Published: January 20, 2024 (Online)

DOI: 10.37936/ecti-cit.2024181.253800

1. INTRODUCTION

The real estate sectors worldwide have overgrown. As such, construction works and subcontractors of relevant business tiers have greatly expanded. Floor and wall tiling is one of the critical elements in a residential building. The quality of such craft is generally assessed by not only its aesthetic aspect but also its functional one. Nonetheless, the scarce supply of skilled craftsmen and experienced inspectors has so far led to numerous defects being found only after the completion of the construction, if not much later. When examining tiling quality, an inspector usually assesses tiling surface, layout patterns, spacing, leveling, and most importantly floor-tile bonding. In fact, faulty bonding between tiles and the underlying floor could result in cracked and flaked off tiles, which often lead to costly reworks.

Generally, to inspect the integrity of floor-tile bonding, the impact-acoustic method is adopted as a non-destructive means of evaluation. This approach involves tapping the tile surface with a small metal object, e.g., a coin or a steel rod. An inspector then carefully listens to the tapping sound and identifies a defect, i.e., whether it is hollow. This type of sound indicates that there may be some problems underneath the tessellation surface. However, acoustic assessment is subjective and depends on the experience, skill, and judgement of individuals. Their proficiency varies by typically uncalibrated, non-standardized, and unreliable equipment, including of course, their hearing.

In the recent research literature, there are many studies that focus on non-destructive methods of debonded tiles inspection. Soeta *et al.* [1] developed a tile-debonding diagnostic device to identify artificially debonded areas of twelve exterior tile specimens. Their device consisted of an impact force sensor and four microphones, each having its output detection parameters. Agarwal [2] exploited millimeter

[1] The author is with the Department of Electrical Engineering, Burapha University, Chon Buri 20131, Thailand., Email: jantanap@eng.buu.ac.th

[2] The author is with the School of Computer Engineering, Suranaree University of Technology, Nakorn Ratchasima 30000, Thailand., Email: phorkaew@sut.ac.th

[1] Corresponding author: jantanap@eng.buu.ac.th

waves for non-invasive crack detection of packaged ceramic tiles using a designed V band (60 GHz) imaging radar system. In addition, Zhao [3] developed a system which inspected wall tile debonding, by using a laser doppler vibrometer and the acoustic waves of sweeping frequency bands. Their system relied on the difference between the activated vibration behavior of the tessellation with and without debonding. Luk [4] proposed another method based on wavelet packet decomposition. Their method allocated the power spectrum density (PSD) of the signal to specific component fields. Their research revealed that the PSD could characterize bonding quality even in noisy environments. Thus, an artificial neural network (ANN) was used as a classifier. Subsequently, they applied a model-based strategy and developed a robust inspection method that utilized wavelet domain features and hidden Markov modeling [5]. Another similar study by Tong [6] inspected tile-wall bonding integrity by analyzing time-domain features of impact acoustic. Firstly, they extracted the feature vectors from the initial acceleration peaks of the impact sound. The features were then carefully selected and automatically classified by a back-propagation ANN. Nonetheless, the performance of the conventional ANN classification model was undermined by the similar characteristics between different bounding, due to abnormal impacts [7]. Instead, the least-squares support vector machine (LS-SVM) was employed in bonding recognition and found to be more reliable and incredibly immune to surface roughness.

Research exists in other different but relevant domains, where machine learning (ML) was used to classify one-dimensional signals. For instance, in biomedicine, electrocardiographic (ECG) signals were analyzed to diagnose cardiovascular diseases [8]-[11]. Toulni [8] processed ECG signals with discrete wavelet transform (DWT). The resultant signals were then classified by an SVM. Kumar [9] proposed a novel method for automatic classification between normal and abnormal signals. In their framework, features were extracted by Fourier analysis and then classified by an SVM. Sun [10] proposed a novel ensemble multi-label classification model, which consisted of various elements, e.g., binary relevance, multi-label k-Nearest Neighbors (k-NN), hierarchical adaptive resonance associative map, twin SVM, sklearn (Scikit-Learn) embedder, and embedding classifier. Similarly, wavelet feature extraction and SVM classification was applied to electroencephalogram (EEG) for diagnosing purpose [11].

In addition to these studies, one-dimensional (1D) signal classification with convolutional neural network (CNN) has also been extensively explored. EEG signals were analyzed for diagnosing brain diseases [12]-[14]. For example, Khan [12] proposed a deep learning (DL) method, called D2PAM, for brain signal classification. The diagnosis of Alzheimer's disease using EEG signal with deep pyramid CNN was proposed in [13]. Moreover, Sandheep [14] applies CNN to classify EEG signals into those of depressed patients and healthy control subjects.

In audio applications [15-20], classification based on ML was also extensively explored. Rong [15] extracted three audio features, i.e., short time energy, zero crossing rate, and Mel-frequency cepstral coefficients (MFCC). These features were then classified by an SVM with Gaussian kernels. Likewise, Lazaro [16] extracted spectral centroid, flatness, and spread, from an audio signal. These features were then classified by SVM to identify the music tempo of the signal. Other features also attracted great interest. For example, Toffa [17] presented a scheme for classifying environmental sounds based on a textural feature encoded by a local binary pattern. Their encoded features were classified with classical ML algorithms, i.e., SVM, random forest (RF), and k-NN. Combining lyrics with audio classification for emotional recognition was also considered [18], where SVM was used. Speech versus music classification based on spectrogram patterns was proposed by Bhattacharjee [19], where they took a novel spectral peak tracking approach.

This study is motivated by a recent investigation [20], in which tile-tapping sounds were classified by CNN, based on their DWT and MFCC features. Albeit found promising, the CNN classifier took considerable time and computing resources. Besides, with relatively small training sets, it also exhibited a tendency of overfitting, even with augmentation. However, the preceding work [20] has laid a foundation that DWT and MFCC features of tapping sound were suitable for the task.

Therefore, this paper aims at strengthening the practical merits of the previous work and to explore various compact but equally effective ML models. Meanwhile, emphasis was placed on systematically configuring and fine-tuning these models to meet on-site criteria. More specifically, it adopts DWT and MFCC feature analyses and classifies them by customized SVM, k-NN, Naïve Bayes, and Logistic Regression algorithms. Their performance metrices were compared with those previously obtained by CNN. It will be demonstrated that the properly configured ML models could perform equally well under resource-limited environments and small labelled samples.

The remainder of this paper is structured as follows. Section 2 explains the characteristics of sounding tile signal and the details of the proposed features (i.e., DWT and MFCC) and classifiers (i.e., SVM, k-NN, Logistic Regression, and Naïve Bayes). The experimental results are presented and discussed in detail in Section 3. Finally, the concluding remarks are given in Section 4.

2. METHODOLOGY

This paper explored various ML configurations, and focused on their effectiveness on tile tapping classification based on MFCC feature extraction [20]. Its main objective was to differentiate between dull and hollow-sounding tiles. The tile tapping signals were acquired at the tiling area at our laboratory. The proposed scheme consisted of three parts, i.e., (1) data preparation, (2) DWT and MFCCs feature extraction, and (3) signal classification based on selected ML models. The scheme is summarized in the following diagram (Fig. 1) and described in detail in the following subsections.

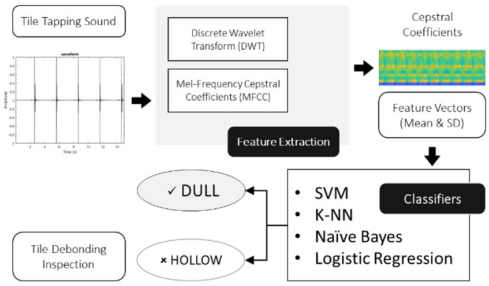


Fig.1: *Overview of the proposed scheme. It consisted of data preparation, feature extraction, and signal classification.*

2.1 Data Preparation

The tapping sound signals were acquired and digitized from the tiling area as shown in Fig. 2(a) [20]. The bonding property of each tile was intentionally formed so that the dull and hollow pieces corresponded to the mapping as shown in Fig. 2(b).

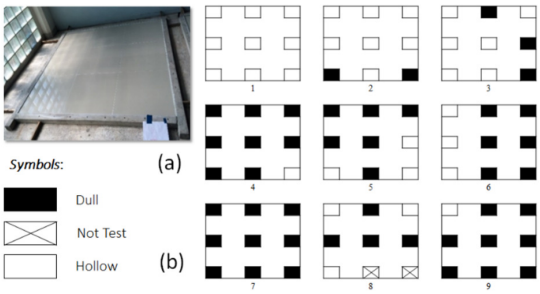


Fig.2: *Tiling area in a laboratory (a) and its bonding map, showing dull and hollow points (i.e., 46 and 33 points, respectively) of each tile.*

Additionally, the tile-tapping device was controlled by an electric motor. The device tapped at nine specified points for each tile at a speed of 20 taps per minute. The resultant 79 audio signals were hence acquired and stored in .wav format. They were divided into two classes, namely (a) dull and (b) hollow sounding. These classes consisted of 46 and 33 signals, respectively [20]. Some examples of dull and hollow sound signals in time and frequency domains are illustrated in Fig. 3 and 4, respectively.

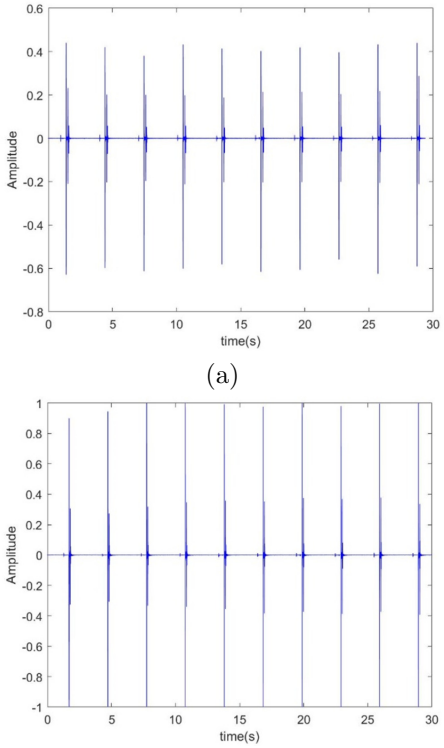


Fig.3: *Selected examples of (a) dull and (b) hollow sounding tiles signals in the time domain.*

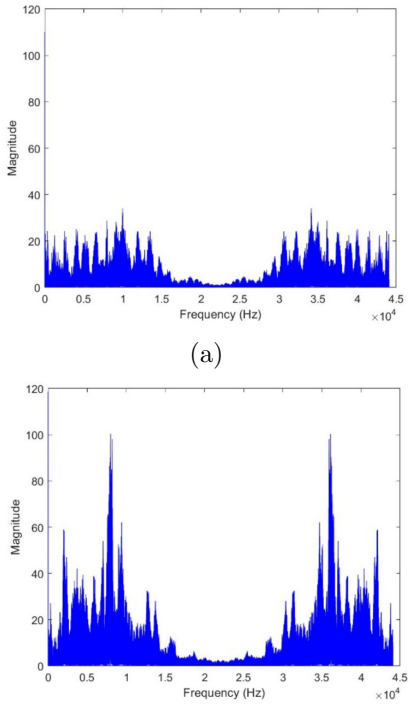


Fig.4: *Selected examples of (a) dull and (b) hollow sounding tiles signals in the frequency domain.*

2.2 Feature Extraction

This study was an extension from a previous work [20]. Therein, it was shown that the 1-level DWT of a tile-tapping acoustic signal combined with its MFCC gave the best classification results. As depicted in Fig. 5, The 1-level DWT decomposition separates an input sound signal into different frequency components.

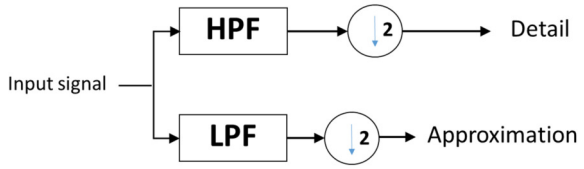


Fig.5: *A 1-level DWT decomposing an input signal via HPF and LPF filters and subsampling the results by a factor of two, giving the detail and approximation parts of the original signal.*

One-level DWT could be computed from an input signal (x) by cascading filters. Specifically, this signal was first sampled and decomposed through a low-pass filter (LPF) with an impulse response (g) and a high-pass filter (HPF) with an impulse response (h). Hence, the filtered signal contained both approximated and detailed coefficients, given, respectively, by low- and high-pass filters. Subsequently, the output signals were obtained by subsampling both filtered signals with a factor of two, as expressed in Eq. (1) and (2), respectively.

$$y_{low}(n) = \sum_{k=-\infty}^{\infty} x(k) \cdot g(2n-k) \qquad (1)$$

$$y_{high}(n) = \sum_{k=-\infty}^{\infty} x(k) \cdot h(2n-k) \qquad (2)$$

In addition to DWT, another feature considered in this study was MFCC. The MFCC extraction consisted of four steps, i.e., (1) Fast Fourier Transform (FFT), (2) Mel-scale filtering, (3) logarithmic scale, and (4) Discrete Cosine Transform (DCT). The detailed analyses and treatments of MFCC are described in [20]. As a result, the filtered spectra were non-linearly mapped on a logarithmic scale. Fig. 6 shows the examples of dull and hollow signal mapping. The amplitudes of the resultant spectra, also called cepstral coefficients, were concatenated to create a feature vector for classification. In the following experiments, higher but less significant orders MFCC coefficients were discarded, leaving only the first thirteen ones, as shown in Fig. 7.

Moreover, to eliminate variations in each MFCC coefficient, the mean and the standard deviation (SD) of cepstral coefficients were calculated for each sample and used as their representatives. In this work, we focused on the two features, i.e., mean value (μ), defined by Eq. (3), and the standard deviation (σ),

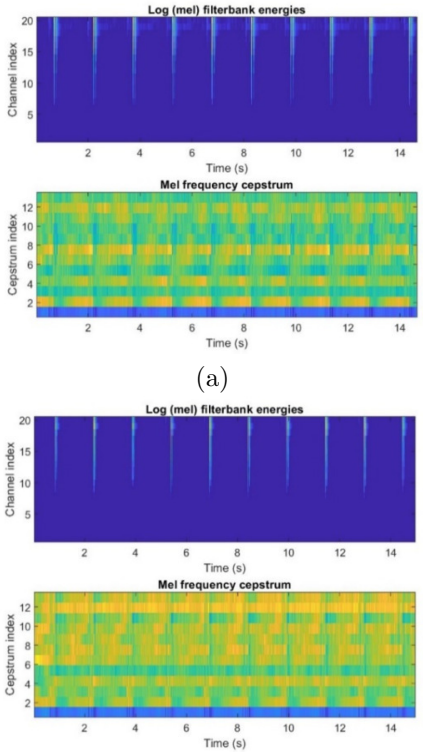


Fig.6: *Log filter bank energy (top) and mel-frequency cepstrum (bottom) of dull (a) and hollow (b) signals.*

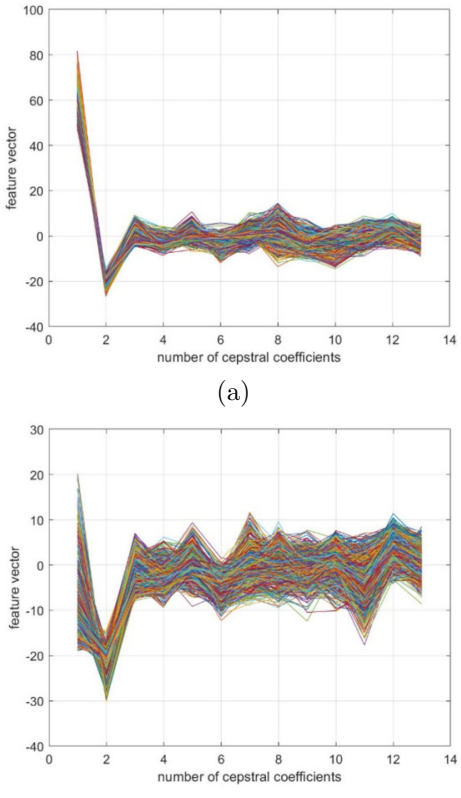


Fig.7: *The first thirteen MFCC coefficients of (a) dull and (b) hollow signals.*

defined by Eq. (4).

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{3}$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \tag{4}$$

where N is the length of a discrete signal $\mathbf{x}$.

2.3 Classification Algorithms

Once extracted from the involved tapping signals, DWT and MFCC were classified by SVM, Logistic Regression, k-NN, and Naive Bayes, in turn. The detailed description of each technique is given in the following subsections.

2.3.1 Support Vector Machine (SVM)

SVM is an ML algorithm typically used in classification, regression, and detection tasks. In this study, it was used to separate features, extracted from audio signals, into different categories. Its main advantages are as follows:

- *Effectiveness:* It can classify data embedded in a multi-dimensional space whose dimensions are more significant than the number of training samples.
- *Usability:* Its decision relies on only a subset of data points called support vectors.
- Multipurpose: An assortment of kernel functions may be specified in the decision function to better suit the problem at hand.

The objective of the SVM is to find a hyperplane in an N-dimensional space (where N is the number of features) that distinctively classifies the data points. In this paper, these SVM terminologies are defined as follows.

a) **Hyperplane:** It is the decision boundary that separates data points, associated with different classes, in an N-D feature space. A linear classifier is defined by a linear equation, i.e.,

$$\mathbf{w^T x} + \mathbf{b} = \mathbf{0} \tag{5}$$

where $\mathbf{w}$ is a normal vector on the hyperplane.
$\mathbf{b}$ is a bias.

b) **Support Vectors (SV)**: They are the closest data points to the hyperplane. These vectors play a vital role in deciding the hyperplane and margin.

c) **Margin**: It is the distance between an SV and the hyperplane. Since the main objective of an SVM is to maximize the margins of all SVs, the more expansive the margins, the better its classification performance.

d) **Kernel**: It is the mathematical function, used to map the original input data points onto a high-dimensional feature space. As a result, the hyperplane can be easily established, even if the data points are not separable in their original space. The standard kernel functions are:

Linear: $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = {\boldsymbol{x}_i}^T \boldsymbol{x}_j$
Polynomial: $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = ({\boldsymbol{x}_i}^T \boldsymbol{x}_j + 1)^q$
Gaussian or radial basis function (RBF): $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2)$
Sigmoid: $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \tanh(\alpha {\boldsymbol{x}_i}^T \boldsymbol{x}_j + b)$

$\mathbf{x_i}$ and $\mathbf{x_j}$ are p-dimensional vectors representing observations i and j, and q is polynomial order.

A binary classification is a subset of SVM models. It labels the two classes with +1 and -1, respectively. Accordingly, we defined a training dataset consisting of input feature vectors $\mathbf{X}$ and their corresponding class labels $\mathbf{Y}$. The distance between any data point $\mathbf{x}$ and the decision boundary can be calculated by:

$$\mathbf{d_i} = (\mathbf{w^T x} + \mathbf{b})/\|\mathbf{w}\| \tag{6}$$

where $\|\mathbf{w}\|$ is the Euclidean norm of the weight vector ($\mathbf{w}$) and $\mathbf{b}$ is a constant vector. Finally, the SVM classifier is given by:

$$\hat{y} = \begin{cases} +1 & , \mathbf{w^T x} + \mathbf{b} \geq \mathbf{0} \\ -1 & , \mathbf{w^T x} + \mathbf{b} < \mathbf{0} \end{cases} \tag{7}$$

2.3.2 Logistic Regression

Logistic regression is a type of statistical model, also named logit model. Generally, it is used for prediction and classification. Logistic regression estimates the probability of an occurring event based on a training dataset of independent variables. The probabilities of these variables are between 0 and 1. A logistic function is represented by a linear regression in Eq. (8) followed by logistic modeling in Eq. (9), i.e.,

$$y = b_0 + \mathbf{b^T x} \tag{8}$$

$$f = 1/(1 + e^y) \tag{9}$$

where $\mathbf{b}$ is the coefficient vector.

2.3.3 k-Nearest Neighbors (k-NN)

The k-NN is a supervised learning classifier that uses proximity to classify (or predict) the group of an individual data point. While the k-NN can solve regression and classification problems, it is typically used for the latter. Its solution is based on the assumption that similar data points are located near one another. Therefore, the goal of k-NN is to identify the nearest neighbors of a queried point. Hence, a lass label is assigned to similar points, belonging to the same group. There are several distance metrics used to measure point proximity. The most typical metrices are as follows.

Euclidean distance is the most frequently used measure, but is normally limited to real-valued vectors.

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\sum_{i=1}^{N} (x_i - y_i)^2} \tag{10}$$

Manhattan distance is another popular distance metric measuring the absolute value between two points.

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{N} |x_i - y_i| \tag{11}$$

Minkowski distance is expressed as a generalized form of Euclidean and Manhattan distance metrics.

$$d(\boldsymbol{x}, \boldsymbol{y}) = \left(\sum_{i=1}^{N} |x_i - y_i| \right)^{\frac{1}{\rho}} \tag{12}$$

where ρ allows for generalization to other distance metrics. Euclidean distance is represented by this formula when $\rho = 2$, while Manhattan distance holds when $\rho = 1$.

2.3.4 Naive Bayes

Naive Bayes is supervised ML. It adopts Bayes' theorem with a naïve assumption on independence condition between every pair of features, given the values of the class variables. Used for classification, but unlike discriminative classifiers (e.g., logistic regression), it does not learn which features are the most important to differentiate between classes. Instead, Bayes' theorem states the following relationship, given the class variable y and dependent feature vectors $\mathbf{x}$, i.e., $\mathbf{x}_\mathrm{i}$ through $\mathbf{x}_\mathrm{n}$, i.e.,

$$P(y|x_1, \ldots, x_n) = \frac{P(y)P(x_1, \ldots, x_n|y)}{P(x_1, \ldots, x_n)} \tag{13}$$

Next, with the naïve conditional independence assumption being imposed, the expression becomes:

$$P(y|x_1, \ldots, x_n) = \frac{P(y) \prod_{i=1}^{N} P(x_i|y)}{P(x_1 P(x_2) \ldots P(x_n)} \tag{14}$$

$\mathrm{P}(\mathbf{x}_1)$,..., $\mathrm{P}(_\mathrm{n})$ are constants given the input. Therefore, the following classification rules may be assumed.

$$P(y|x_1, \ldots, x_n) \propto P(y) \prod_{i=1}^{N} P(x_i|y) \tag{15}$$

$$y = \arg\max_y (P(y) \prod_{i=1}^{N} P(x_i|y)) \tag{16}$$

This study used the Maximum A Posteriori (MAP) strategy to estimate P(y) and $\mathrm{P}(\mathbf{x}_\mathrm{i} \parallel y)$.

3. RESULTS AND DISCUSSION

This section describes the experimental settings and the performance evaluation of the proposed classification. In particular, the classification results obtained by the involved ML algorithms were compared with those from the previous study [20]. Finally, the last subsection provides the relevant discussions on the comparisons.

3.1 Objective Evaluation

In this subsection, the following metrices were evaluated for each ML.

Accuracy (Acc.) of a classification was expressed by the ratio between the number of correctly classified signals and that of all signals. It was given by Eq. (17).

$$Acc. = \frac{n}{N} \tag{17}$$

where n is the number of correctly classified signal samples, and N is the total ones.

Specificity, also known as true negative rate, measures the proportion of actual negatives. In this study, it was defined as the percentage of hollow-sounding tiles that were correctly identified. It was given in Eq. (18).

$$Spec. = \frac{TN}{TN + FP} \tag{18}$$

where TN is true negative, and FP is false positive.

Sensitivity, also known as true positive rate, measures the proportion of actual positives. Likewise, it was defined as the percentage of correctly identified dull-sounding tiles. Its expression is given in Eq. (19).

$$Sen. = \frac{TP}{TP + FN} \tag{19}$$

where TP is true positive, and FN is false negative.

Receiver Operating Characteristic (ROC) curve is a graph showing the performance of a classification model. The curve plots between true positive rate (TPR) and false positive rate (FPR). While the former is defined the same as the sensitivity in Eq. (19), the latter is given in Eq. (20).

$$FPR = \frac{FP}{FP + TN} \tag{20}$$

AUC, also known as area under the *ROC* curve, measures the entire two-dimensional area under the *ROC* curve from (0, 0) to (1, 1).

3.2 Experimental Results

The tapping signals were recorded at the tiling area, as shown in Fig. 2(a). Since only limited samples were tested, the Leave-one-out cross-validation was performed to refute model overfitting. The performance of different classifiers (i.e., k-NN, SVM, Logistic Regression, and Naive Bayes), was evaluated, based on the standard metrices (i.e., accuracy, sensitivity and specificity).

Firstly, the k-NN classifier was evaluated by adjusting the number of nearest neighbors (#NN) from

2 to 8. The results are reported in Table 1 and Fig. 8 and 9. It is evident that with #NN = 4 and #NN = 8, the accuracy was the highest at 88.6076 %. Likewise, the specificity was the highest when #NN = 8, while the sensitivity was the highest when #NN = 4.

Table 1: *Evaluation metrics obtained by the k-NN.*

#NN	Acc (%)	Spec.	Sen.	AUC
2	84.8101	0.8387	0.8542	0.9941
3	84.8101	0.9200	0.8148	0.9842
4	**88.6076**	0.9000	**0.8776**	0.9743
5	86.0759	0.9583	0.8182	0.9634
6	87.3418	0.9259	0.8462	0.9664
7	84.8101	0.9565	0.8036	0.9694
8	**88.6076**	**0.9615**	0.8491	0.9638

Note #NN is the number of nearest neighbors.

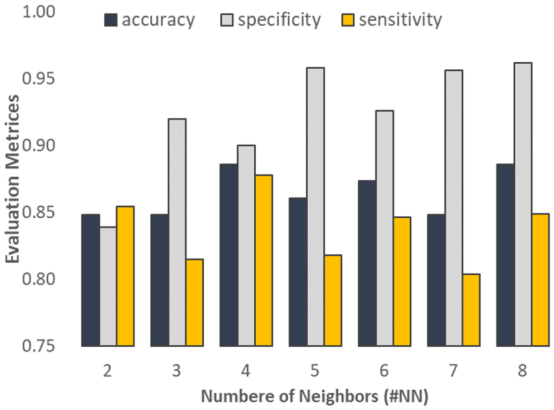


Fig.8: *Comparison of accuracy, specificity, and sensitivity using k-NN.*

Since the scores were tied, the one that required the least number of nearest neighbors (i.e., faster to compute) was considered optimal. In addition, it is worth remarking that increasing this number did not significantly affect the overall performance of the system.

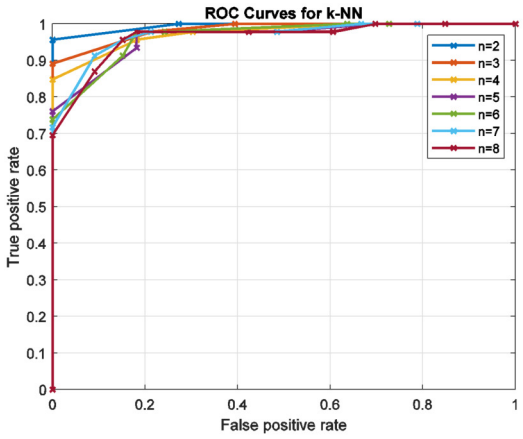


Fig.9: *ROC Curve for k-NN classifier.*

Then, the SVM was evaluated by taking linear, Gaussian, and polynomial kernels in turns. The results reported in Table 2 and Fig. 10 indicate that linear SVM performed best in all metrics for our 2-class problem.

Table 2: *Evaluation metrics obtained by SVM classifier with different kernels.*

Kernel	Acc. (%)	Spec.	Sen.	AUC
Linear	**84.8101**	**0.8387**	**0.8542**	0.9743
Gaussian	58.2278	–	0.5823	1
Polynomial	34.1772	0.3699	0	0.9565

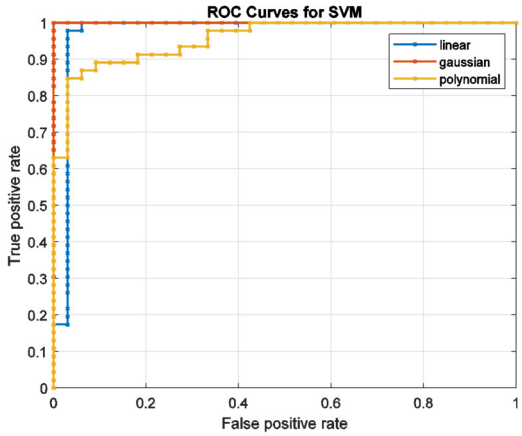


Fig.10: *ROC Curve for SVM classifier.*

Another tunable element in a given SVM is the solver optimization routine. In the next experiment, linear SVM was specified. In each round, either sequential minimal optimization (SMO), iterative single data (ISDA), or L1 soft-margin minimization by quadratic programming (L1QP) routine was employed, in turn. It was thus revealed in Table 3 that all classifications gave the same outcomes, regardless of the optimizing routines.

The suggests that extracted features and the classifier were invariant against internal implementation and thus suitable for the task. It is worth pointing out that, on the contrary, selecting different loss functions and optimizers in a CNN-based DL could vastly affect its performance [21-22].

Table 3: *Evaluation metrics obtained by Linear SVM classifier with different solvers.*

Solver	Acc (%)	Spec	Sen
SMO	84.8101	0.8387	0.8542
ISDA	84.8101	0.8387	0.8542
L1QP	84.8101	0.8387	0.8542

Finally, automated kernel scaling was activated for all kernels. The results remained consistent with those reported in Table 4 and Fig. 11. However, its accuracy was raised to 93.6709 % for the linear kernel. This result suggests that auto kernel scaling is preferred in the current settings.

Table 4: *Evaluation results obtained by linear SVM classifier with auto kernel scaling enabled.*

Kernel	Acc. (%)	Spec.	Sen.	AUC
Linear	**93.6709**	**0.9375**	**0.9362**	0.9480
Gaussian	89.8734	0.8788	0.9130	0.9974
Polynomial	89.8734	0.8788	0.9130	1

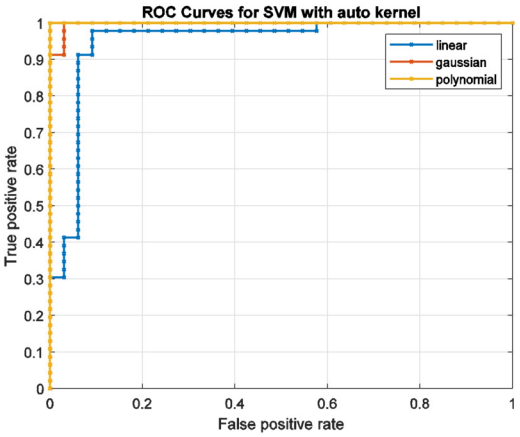
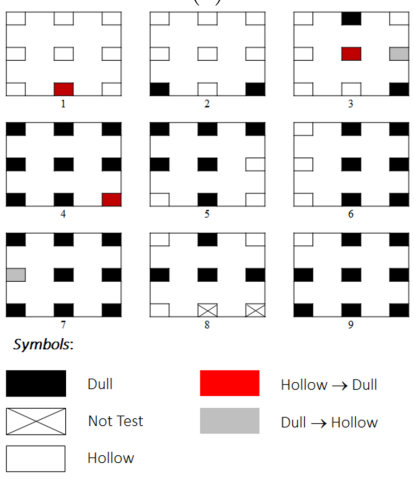


Fig.11: *ROC Curve for SVM classifier with auto kernel scaling enabled.*

Accordingly, the resultant confusion metric of the most accurate classification is illustrated in Fig. 12(a) with the corresponding tile mapping in Fig. 12(b). In Fig. 12(b), red and grey points indicate incorrect detections (i.e., false positives and false negatives, respectively). Black and white points indicate correct detections (i.e., true positives and true negatives, respectively).

		Actual class	
		Dull	Hollow
Predicted class	Dull	44	2
	Hollow	3	30

(a)



(b)

Fig.12: *Confusion matrix of the optimal SVM classifier (a) and the corresponding tile prediction map (b).*

Likewise, the next classifier to be evaluated was the Naïve Bayes model. In the experiment, data distribution was varied among Gaussian (i.e., normal), kernel smoothing density estimate (i.e., kernel), and multivariate multi-nominal (i.e., mvmn) distributions. The results reported in Table 5 and Fig. 13 suggest the first distribution gave the most accurate classification at 91.1392 %.

Table 5: *Evaluation results obtained by Naïve Bayes classifier with varied distribution types, i.e., normal, kernel, and mvmn.*

Distribution	Acc (%)	Spec	Sen	AUC
normal	**91.1392**	**0.8824**	0.9333	0.9598
kernel	88.6076	0.8333	0.9302	0.9895
mvmn	58.2278	–	0.5823	1

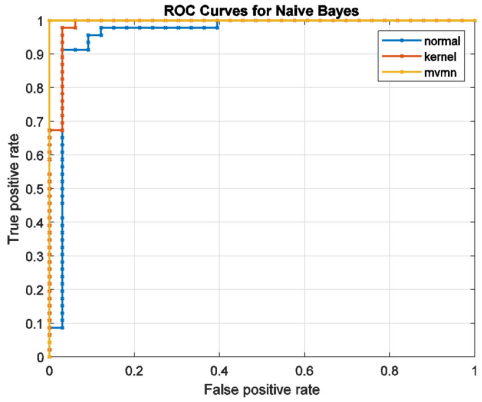


Fig.13: *ROC curve for Naïve Bayes classifier.*

When comparing the basic linear SVM with logistic regression, it is evident in Fig. 14 that the latter performed marginally better in evaluation metrics. Then, the optimized Naïve Bayes was examined. The optimization took a longer time to converge. However, it reached the same level of accuracy as the optimized (kernel) linear SVM, as reported in Table 4. Since the evaluation metrices were exactly the same, they are not repeated here. It is, however, safe to conclude that both classifiers yielded the best classification results.

Fig.14: *Comparison of results obtained by basic linear SVM and Logistic Regression.*

They were as accurate as when using more advanced classifiers, like CNN [20], but incurred much fewer computing resources and processing time. In addition, Table 6 and Fig. 15 compare the accuracy (Acc.) in %, specificity (Spec.) and sensitivity (Sen.) between the employed ML and the CNN classifications.

Table 6: *Comparison of Machine Learning and Deep Learning.*

Methods	Acc. (%)	Spec.	Sen.
k-NN with 8 NN	88.6076	0.9615	0.8491
SVM linear kernel	84.8101	0.8387	0.8542
SVM with auto linear kernel scaling	**93.6709**	**0.9375**	**0.9362**
Naïve Bayes with normal distribution	91.1392	0.8824	0.9333
CNN [20]	**93.6709**	**0.9375**	**0.9362**

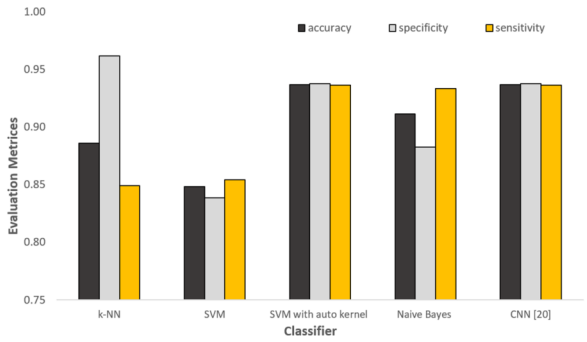


Fig.15: *Comparison of the proposed method and ref [20].*

4. CONCLUSION

This paper proposes a novel cost-effective computerized tapping sound classification for tile inspection. The method was inspired by the recent success of a similar classification, based on DWT and MFCC features [20]. In addition, various prominent ML models, i.e., k-NN, SVM, Naïve Bayes, and Logistic Regression were examined as the classifiers. The experimental results indicated that the developed and fine-tuned system was able to distinguish dull from hollow sounding tiles at 93.67 % accuracy. The specificity and sensitivity were 93.75 % and 93.62 %, respectively. Among the involved models, the optimized SVM and Naïve Bayes ones performed best for the task.

Herein, it is demonstrated that model performance was invariant of trivial configurations other than kernel type. Therefore, it is reasonable to anticipate that the proposed classification system can serve as a guideline for standardizing the automated tiling inspection in practice as off-the-shelf software.

Future research directions worth considering include investigating model variations given different tile dimensions, shapes, thicknesses, and materials. Moreover, experiments comparing various combinations of features and ML models against the DWT and MFCC would fill the research gaps present in this study.

ACKNOWLEDGEMENT

The authors are thankful to Dr. Petcharat Limsupreeyarat for the tile-tapping sound and the anonymous reviewers whose constructive critiques helped improve the quality of this paper.

References

[1] T. Soeta, S. Ito, T. Fujinuma and T. Mikami, "Trial Inspection of Exterior-tile Wall Specimens with a Prototyped Tile-debonding Diagnostic Device," *Japan Architectural Review*, vol. 4, no. 1, pp. 28-40, 2021.

[2] S. Agarwal and D. Singh, "An Adaptive Statistical Approach for Non-Destructive Underline Crack Detection of Ceramic Tiles Using Millimeter Wave Imaging Radar for Industrial Application," *IEEE Sensors Journal*, vol. 15, no. 12, pp. 7036-7044, 2015.

[3] Y. Zhao, Y. Chen and L. Ye, "A Non-contact Inspection Method of Tile Debonding using Tuned Acoustic Wave and Laser Doppler Vibrometer," *Journal of Sound and Vibration*, vol. 117875, pp. 1-15, 2023.

[4] B.L. Luk a, K.P. Liu, Z.D. Jiang, and F. Tong, "Robotic Impact-acoustics System for Tile-wall Bonding Integrity Inspection," *Mechatronics*, vol. 19, no. 8, pp. 1251-1260, 2009.

[5] B.L. Luk, K.P. Liu, F. Tong, and K.F. Man, "Impact-Acoustics Inspection of Tile-wall Bonding Integrity via Wavelet Transform and Hidden Markov Models," *Journal of Sound and Vibration*, vol. 329, no. 10, pp. 1954-1967, 2010.

[6] F. Tong, S.K. Tso and X.M. Xu, "Tile-wall Bonding Integrity Inspection based on Time-domain Features of Impact Acoustics," *Sensors and Actuators A: Physical*, vol. 132, no. 2, pp. 557-566, 2006.

[7] F. Tong, X.M. XU, B.L. Luk and K.P. Liu, "Evaluation of Tile–wall Bonding Integrity based on Impact Acoustics and Support Vector Machine," *Sensors and Actuators A: Physical*, vol. 144, no. 1, pp. 97-104, 2008.

[8] Y. Toulni, N. Benayad and B. D. Taoufiq, "Electrocardiogram Signals Classification using Discrete Wavelet Transform and Support Vector Machine Classifier," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 2, pp. 960-970, 2021.

[9] A. Kumar, V. K. Mehla, H. Tomar, M. Kumar and R. Komaragiri, "Classification of Nor-

mal and Abnormal ECG Signals using Support Vector Machine and Fourier Decomposition Method," *Proceeding of International Symposium on Smart Electronic System*, pp. 161-166, 2020.

[10] Z. Sun, C. Wang, Y. Zhao and C. Yan, "Multi-label ECG Signal Classification based on Ensemble Classifier," *IEEE Access*, vol. 8, pp. 117986-117996, 2020.

[11] N. T. Br. Pasaribu, T. Halim, Ratnadewi and A. Prijono, "EEG Signal Classification for Drowsiness Detection using Wavelet Transform and Support Vector Machine," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 2, pp. 501-509, 2021.

[12] A. A. Khan, R.K. Madendran, U. Thirunavukkaras and M. Faheem, "D2PAM: Epileptic Seizures Prediction using Adversarial Deep Dual Patch Attention Mechanism," *CAAI Transactions on Intelligence Technology*, vol. 8, no. 3, pp. 755-769, 2023.

[13] W. Xia, R. Zhang, X. Zhang and M. Usman, "A Novel Method for Diagnosing Alzheimer's Disease using Deep Pyramid CNN based on EEG Signals," *Heliyon*, vol. 9, No. 4, 2023.

[14] P. Sandheep, S. Vineeth, M. Poulose and D. P. Subha, "Performance Analysis of Deep Learning CNN in Classification of Depression EEG Signals," *Proceeding of IEEE Region 10 Conference (TENCON)*, pp. 1339-1344, 2019.

[15] F. Rong, "Audio Classification Method based on Machine Learning," *Proceeding of International Conference on Intelligent Transportation, Big Data & Smart City*, pp. 81-84, 2016.

[16] A. Lazaro, R. Sarno, R. J. Andre, and M. N. Mahardika, "Music Tempo Classification using Audio Spectrum Centroid, Audio Spectrum Flatness, and Audio Spectrum Spread based on MPEG-7 Audio Features," *Proceeding of 3rd International Conference on Science in Information Technology*, pp. 41-46, 2017.

[17] O. K. Toffa and M. Mignotte, "Environmental Sound Classification using Local Binary Pattern and Audio Features Collaboration," *IEEE Transactions on Multimedia*, vol. 23, pp. 3978–3985, 2020.

[18] W. Shi and S. Feng, "Research on Music Emotion Classification Based on Lyrics and Audio," *Proceeding of Advanced Information Technology, Electronic and Automation Control Conference*, pp. 1154-1159, 2018.

[19] M. Bhattacharjee, S. R. M. Prasanna and P. Guha, "Speech/Music Classification using Features from Spectral Peaks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 1549-1559, 2020.

[20] J. Panyavaraporn, P. Limsupreeyarat and P. Horkaew, "DWT/MFCC Feature Extraction for Tile Tapping Sound Classification," *International Journal of Integrated Engineering*, vol. 12, no. 3, pp. 122-130, 2020.

[21] E. M. Dogo, O. J. Afolabi and B. Twala, "On the Relative Impact of Optimizers on Convolutional Neural Networks with Varying Depth and Width for Image Classification," *Applied Sciences*, vol. 12, no. 23: 11976, 2022.

[22] Vidushi, M. Agarwal, A. Rajak and A.K. Shrivastava, "Assessment of Optimizers impact on Image Recognition with Convolutional Neural Network to Adversarial Datasets," *Journal of Physics: Conference Series*, vol. 1998, no. 1: 012008, 2021.

Jantana Panyavaraporn received the B.Eng. in Electrical Engineering from Burapha University (2002), M.Eng. in Telecommunication Engineering from King Mongkut's Institute of Technology Ladkrabang (2005) and Ph.D. in Electrical Engineering from Chulalongkorn University (2010). She is an Associate Professor at Electrical Engineering, Burapha University, Thailand. Her research interests include image processing, video processing, video coding and computer vision.

Paramate Horkaew received the B.Eng. (Hons.) degree in telecommunication engineering (1st) from the King Mongkut's Institute of Technology Ladkrabang, Ladkrabang, Thailand, in 1999, and the Ph.D. degree in computer science from the Imperial College London, U.K., in 2004. He is currently an Associate Professor at the School of Computer Engineering, Suranaree University of Technology, Thailand. His main research interests include remote sensing, medical imaging, digital geometry processing, pattern recognition, computer vision and graphics.