Electronic supplementary materials

For https://doi.org/10.1631/jzus.A2500545

Novel robust cross-modal integration fusion model for rapid moisture content detection in concrete sand

Zhijian CAI, Jun ZHANG, Xiaoling WANG, Jiajun WANG, Kehao ZHAO, Guohua WU

State Key Laboratory of Hydraulic Engineering Intelligent Construction and Operation, Tianjin University, Tianjin 300350, China

Section S1

After initializing the MC of all sand samples, the multimodal data of the NIR spectra, images, and DC of the sand were simultaneously collected at 30-minute intervals. The NIR data of the sand were collected using the NIR-310L (Shenzhen Pynect Science and Technology, Co., China). The device has a spectral band range of 1350–2150 nm and a resolution of 5 nm. The spectrometer was calibrated using a standard whiteboard prior to the measurements. After calibration, the measuring hole of the spectrometer was directly attached to the surface of the sand material in the aluminum box. Six spectral datapoints were continuously measured at six randomly selected points for each sample, and the average value was taken as the representative spectrum of the corresponding MC. Following the NIR data acquisition, the sample was placed in a holder that had previously secured an iPhone 13 device (Apple Inc.). To reduce the interference of ambient light on the sample images, the flash photography method was adopted to collect sand images with sizes of 4032 px × 3024 px. After obtaining the image data, an FDR probe (TRS-1203, manufactured by Beijing ShiYang Electronics Technology Co., Ltd.) was fully inserted into the sample. Six positions were randomly selected for measurements, and the average of the six measurements was taken as the DC of the sand at the corresponding MC. The entire multimodal data acquisition process took approximately 70 s, which greatly improved detection efficiency compared with the traditional drying method. Following the data collection of the three modalities, the samples were placed in the laboratory at 25 ℃ to simulate natural volatilization. Subsequently, the samples were uniformly stirred from the bottom up for approximately 1 minute at 30-minute intervals, and the above procedure was repeated. Upon pausing the experiment, the samples were sealed using a new impermeable membrane to facilitate subsequent multimodal data sampling.

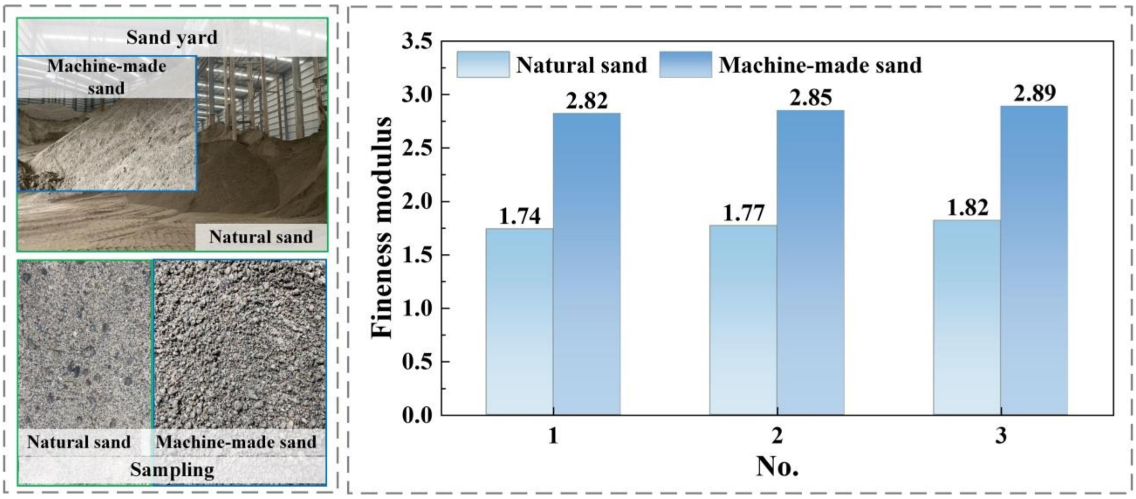

Fig. S1. Sand sample fineness modulus.

Table S1. Statistical characteristics of the MC of the selected concrete sand.

Sand type	Property	Dataset	Max	Min	Mean	Median	Std	Number
Machine-made	MC (wt.-%)	ALL	14.0	0.	7.01	6.68	13.5	147
		Training set	14.0	0.	7.01	6.68	13.9	99
		Validation set	13.4	1.	7.09	6.70	13.1	24
		Test set	13.2	0.	6.91	6.54	13.1	24
Natural	MC (wt.-%)	ALL	20.5	0	9.29	8.66	32.0	183
		Training set	20.5	0	9.29	8.66	32.7	123
		Validation set	19.9	0.	9.38	8.79	31.8	30
		Test set	19.2	0.	9.16	8.67	31.3	30

Section S2 Methodology

Section S2.1 NIR feature extraction under ICNN and AT-BiGRU framework

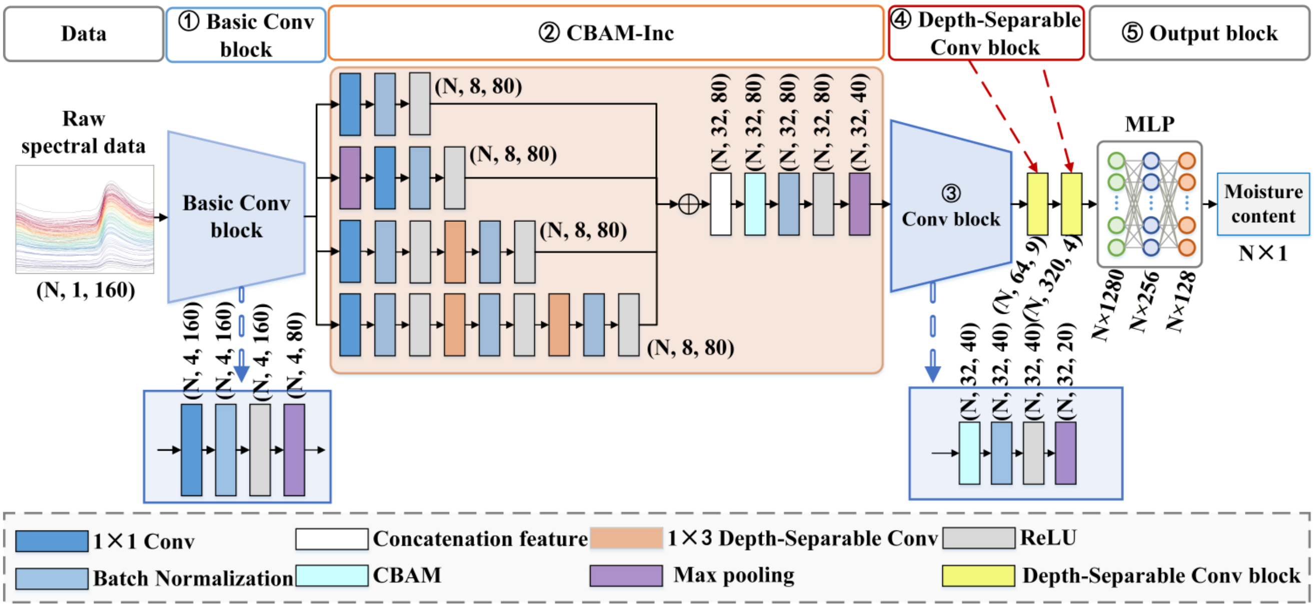

Fig. S2. Flowchart of ICNN.

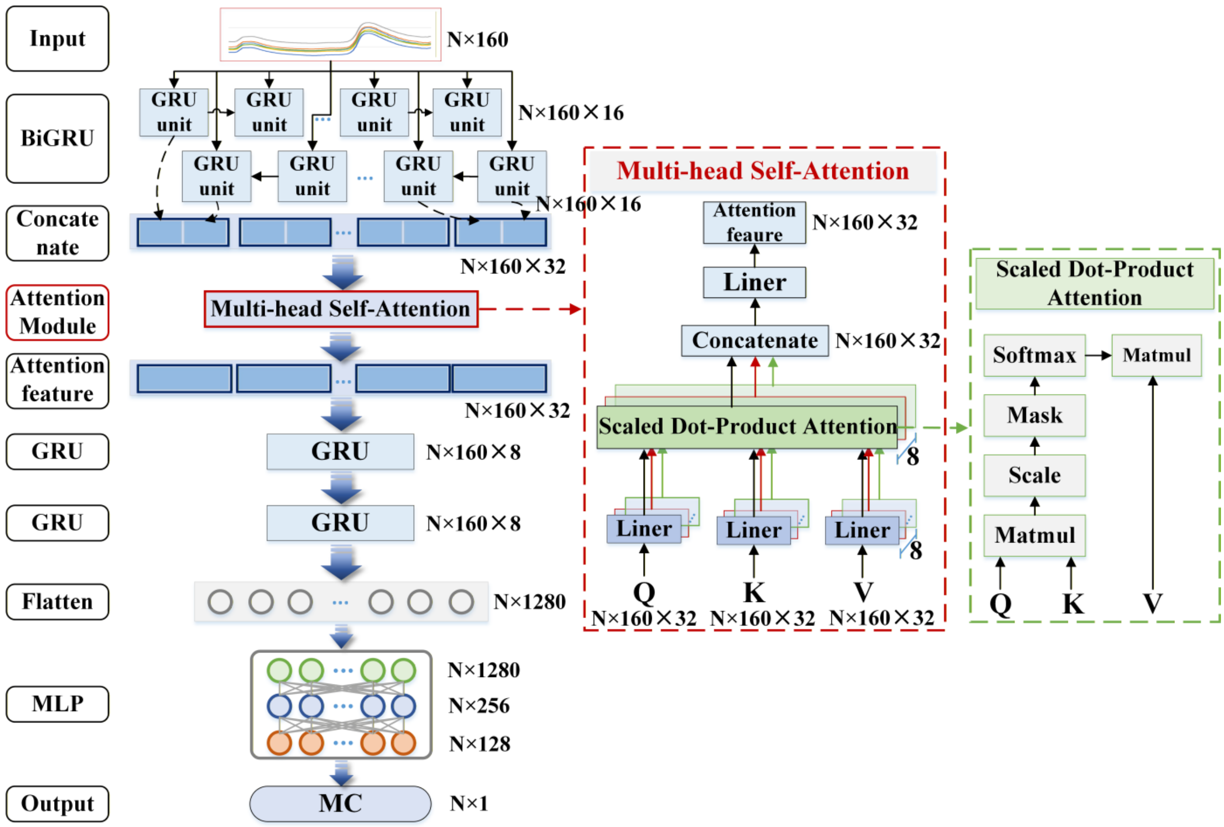

Fig. S3. Flowchart of AT-BiGRU.

Section S2.2 Image feature extraction under improved EfficientNet

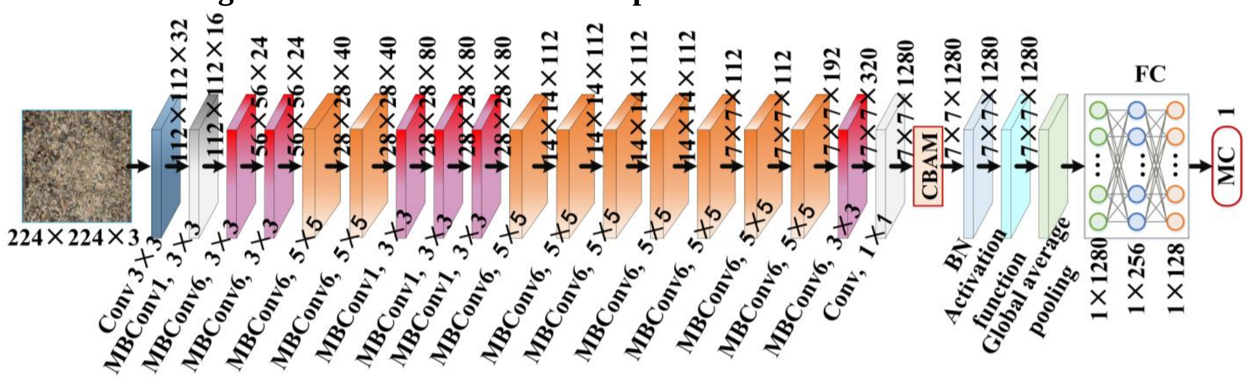

Fig. S4. Flowchart of EfficientNet.

Section S2.3 DC feature extraction under ANN and GWOSVR

$$f(x) = w^T \bullet \psi(x_i) + b \tag{S1}$$

where, $f(x)$ is the regression hyperplane; w is the weight vector; $\psi(x_i)$ and b are the weight coefficient and bias, respectively. $f(x)$ can be equated to an optimization solution question as follows:

$$\min R(w,b,\xi) = \frac{1}{2}\|w\|^2 + c\sum_{i=1}^{N}(\xi_i + \xi_i^*) \tag{S2}$$

where, c is the penalty, ξ_i and ξ_i^* are the relaxation variables.

Section S3 Results and analysis

Section S3.1 Experimental results for concrete sand

Fig. S5 presents the performance of the proposed NID-MCIF model on the training, validation, and test sets. It can be observed that the model achieves high accuracy across all three sets for both sand types. Furthermore, the differences in R^2 between the training set and the validation and test sets are marginal (< 0.01), with only a limited increase in RMSE, thereby verifying the strong generalization capability of the proposed model.

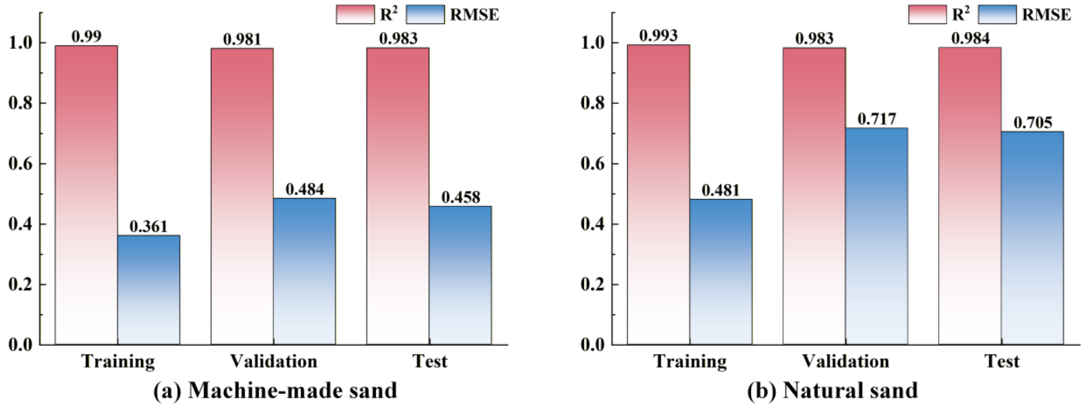

Fig. S5. Training, Validation, and Test performance statistics of the NID-MCIF model.

Section S3.2 Analysis of multimodal fusion weights

An interpretability analysis of the coupling relationship between the multimodal data of the proposed model during the detection of the MC in sand was performed, as shown in Fig. S6. In the process of NIR abstract feature fusion (NIR1_fusion) of machine-made sand, the NIR positional features (NIR2) exhibit the largest weight (0.5). The weights of image features (IMG) and NIR abstract features (NIR1) are 0.313 and 0.187, respectively. This indicates that the NIR2 and IMG features contain more critical information. This can be attributed to the relatively smooth nature of NIR data and its weak nonlinear variations, leading to suboptimal extraction of abstract features. Moreover, NIR positional features exhibit strong positional correlations, reflecting the superiority of AT-BiGRU in capturing long-range dependencies in sequential features. For machine-made sand, the material surface exhibits varying textures and color information at different moisture contents, providing an abundance of valuable features. During the fusion of NIR positional features and image features, the NIR positional features still hold the highest weight (0.349). This can be explained by the fact that NIR positional features provide more key information, while the relatively weaker NIR abstract and image features are able to acquire the optimal enhanced features through weighted fusion with comparable weights. In natural sand, during the fusion process of multimodal data, NIR positional features still carry the largest weight, while NIR abstract features and image features acquire the optimal enhanced features through weighted fusion with comparable weights, following the same pattern as observed in manufactured sand. This demonstrates that the proposed cross-modal self-attention fusion module can dynamically assign weights to different modal features and establish the relationships between them, thereby facilitating the extraction of optimal enhanced features.

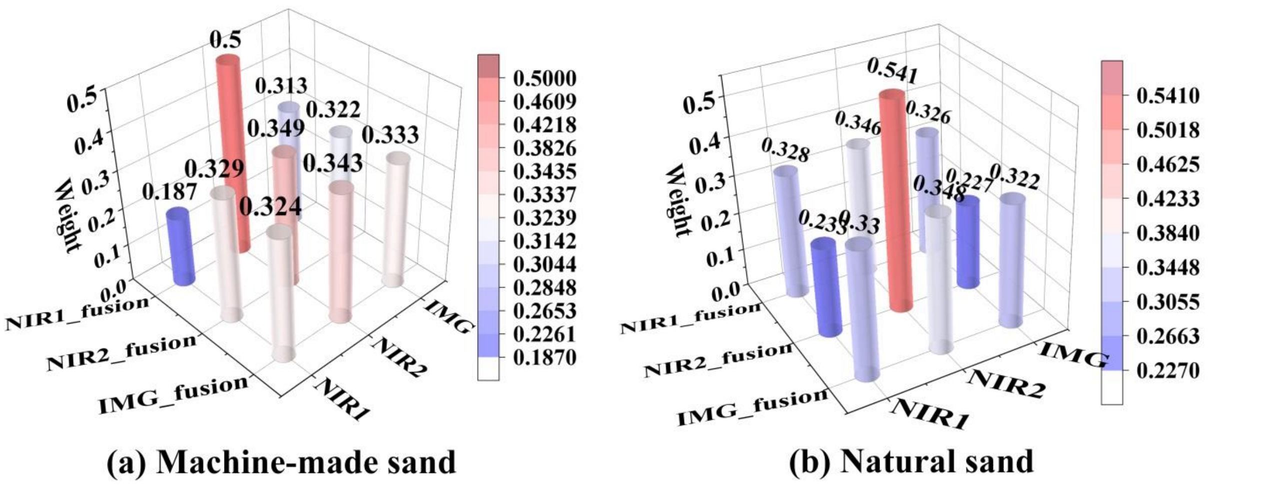

Fig. S6. Cross-modal fusion weight distribution of the proposed model.

The weights assigned to the multi-level, multi-modal fusion features within the integrated output module, with respect to the final output, are presented in Table S2. The weights of the machine-made sand MC outputs are sorted as follows: $Pred_{DC}(0.749)$ > $Pred_{Img}^{fusion}(0.401)$ > $Pred_{NIR1}^{fusion}(0.258)$ > $Pred_{NIR2}^{fusion}(0.181)$ > $Pred_{ALL}^{fusion}(0.081)$. The weights of the natural sand MC outputs are ranked as follows: $Pred_{DC}(0.744) > Pred_{NIR2}^{fusion}(0.714) > Pred_{NIR1}^{fusion}(0.68) > Pred_{ALL}^{fusion}(0.586) > Pred_{Img}^{fusion}(0.522)$. The above results indicate some differences in the weight of the MC output of each modal fusion feature in different datasets. The contribution of multi-modal fusion features is not constant but varies significantly with the type of sand material. The DC feature has the greatest impact, as its probe is fully inserted into the material, resulting in relatively less interference from environmental factors. Image fusion features exhibit varying sensitivities to MC changes in texture and color, depending on the material type. NIR abstract and positional fusion features differ due to variations in environmental lighting and collection locations. It is also demonstrated that the multi-level cross-modal integration fusion network can effectively integrate the advantages of multi-level fused features, achieving robust MC detection results through dynamic weighting.

Table S2. Weight of the prediction results of each modal fusion feature in the integrated output module.

Sand type	Weight				
	$Pred_{NIR1}^{fusion}$	$Pred_{NIR2}^{fusion}$	$Pred_{Img}^{fusion}$	$Pred_{ALL}^{fusion}$	$Pred_{DC}$
Machine-made sand	0.258	0.181	0.401	0.081	0.749
Natural sand	0.68	0.714	0.522	0.586	0.744

Section S3.3 Ablation experiment

The ablation experiments on the proposed model were conducted to verify the effectiveness of the MCIF proposed in this study (refer to Table S3). This study systematically investigates the performance impact of various module combinations within the MCIF framework on the proposed MCIF-based trimodal fusion prediction model, with corresponding statistical characteristics detailed in Table S3. Compared with NID-MCIF_without_A、NID-MCIF_without_B, and NID-MCIF_without_C, the NID-MCIF exhibits IR of {9.7%, 25.4%, 28.8%} and {22.9%, 17%, 19.34%} in the MC detection results for machine-made and natural sands, respectively. The results indicate that the absence of the module A had a relatively minor impact on machine-made sand but a more pronounced effect on natural sand. This can be explained by variations in environmental conditions such as lighting, temperature, and sample differences during the experiments, which introduced varying levels of noise, leading to differences in the effectiveness of feature attention across sand types. The separate absence of module B and module C induces significant performance degradation for both machine-made sand and natural sand, with the absence of module C causing more severe degradation than module B. These results substantiate that considering potential inter-modality interactions effectively enhances model prediction performance, while integrating multi-level feature outputs demonstrates greater potential for predictive capability improvement. Compared with NID-MCIF_without_AB、NID-MCIF_without_AC, and NID-MCIF_without_BC, the NID-MCIF exhibits IR of {38.9%, 37.2%, 39.6%} and {26.3%, 29.9%, 41.7%} in the MC detection results for machine-made and natural sands, respectively. The combined deficiency of modules B and C exerts the most significant impact on both machine-made and natural sand MC detection performance. For machine-made sand, the joint absence of modules A and B demonstrates comparable effects on model performance to that of modules A and C, whereas for natural sand, the former exhibits relatively smaller influence. Synthesizing these findings reveals that module C exerts the most substantial effect on model performance, followed by module B, with module A showing the least impact, underscoring that the multi-level feature integration fusion strategy is crucial for significantly improving the model's detection accuracy. Compared with NID-CC, the NID-MCIF exhibits IR of 48.2% and 44.9% in the MC detection results for machine-made and natural sands, respectively. The results indicate that the proposed model achieved the most significant performance improvement compared to the model using direct multimodal feature concatenation. Additionally, it demonstrated high accuracy in MC detection across different types of sand materials. This highlights that the proposed MCIF effectively enhances MC detection accuracy and stability by accounting for its incorporation of latent intra- and inter-modality interactions coupled with fusion of multi-level feature outputs.

The comparative analysis of model parameters and floating-point operations (FLOPs) indicates that the parameters of the proposed NID-MCIF model are merely 1.38, 1.99, and 1.03 times those of the ablated models (i.e., models with module A, B, or C removed, respectively), while their FLOP counts remain comparable. However, the proposed model exhibits a significant performance enhancement, with IR ranging from a minimum of 9.7% to a maximum of 48.2% over the comparison models. This verifies that the performance gain of the NID-MCIF model primarily originates from the synergistic design among its modules, rather than a mere increase in parameter volume.

Table S3. Comparison of the performance of the proposed model before and after improvement.

Model	Machine-made sand			Natural sand			Params	FLOPs
	R^2	RMSE	RPD	R^2	RMSE	RPD		
NID-MCIF	**0.983**	**0.458**	**7.9**	**0.984**	**0.705**	**7.931**	**39.633M**	**3.377G**
NID-MCIF_without_A	0.979	0.507	7.133	0.972	0.914	6.120	28.741M	3.356G
NID-MCIF_without_B	9.700	0.614	5.898	0.976	0.849	6.591	19.963M	3.337G
NID-MCIF_without_C	0.969	0.643	5.631	0.975	0.874	6.398	38.544M	3.375G
NID-MCIF_without_AB	0.958	0.750	3.531	0.969	0.957	5.843	10.128M	3.318G
NID-MCIF_without_AC	0.959	0.729	3.709	0.967	1.005	5.565	28.710M	3.355G
NID-MCIF_without_BC	0.954	0.759	4.770	0.953	1.210	4.622	18.874M	3.335G
NID-CC	0.938	0.884	4.097	0.945	1.279	4.372	9.066M	3.315G

Section S4 Discussion

Section S4.1 Comparison with the unimodal model

Table S4. Performance statistics of the proposed model and unimodal model.

Model	Machine-made sand			Natural sand			Params	FLOPs
	R^2	RMSE	RPD	R^2	RMSE	RPD		
NID-MCIF	**0.983**	**0.458**	**7.9**	**0.984**	**0.705**	**7.931**	**39.633M**	**3.377G**
ICNN (NIR)	0.868	1.288	2.810	0.917	1.588	3.523	0.362M	0.860M
AT-BiGRU (NIR)	0.893	1.158	3.126	0.912	1.63	3.431	0.364M	1.873M
CNN-GRU (NIR)	0.886	1.197	3.023	0.901	1.733	3.228	0.719M	4.487M
CNN-GRU-MCIF	0.903	1.102	3.286	0.935	1.407	3.976	21.090M	44.803M
CNN-GRU-DMF	0.907	1.078	3.358	0.929	1.464	3.823	0.131M	2.886M
CNN-GRU-SAF	0.894	1.153	3.140	0.920	1.558	3.591	13.812M	30.254M
GWOSVR (NIR)	0.619	2.186	1.656	0.804	2.433	2.3	8836	26220
ANN (NIR)	0.849	1.378	2.627	0.896	1.775	3.152	74241	0.148M

Model	Machine-made sand			Natural sand			Params	FLOPs
	R^2	RMSE	RPD	R^2	RMSE	RPD		
GWOSVR (DC)	0.952	0.774	4.679	0.957	1.147	4.879	981	2450
ANN (DC)	0.892	1.165	3.106	0.936	1.39	4.024	151	250
IEfficientNet (Image)	0.943	0.843	4.293	0.957	1.143	4.893	4.573M	0.829G

Section S4.2 Comparison with the multimodal model

This study further investigates different modality combinations under the same fusion strategy. As shown in Table S5, the proposed NID-MCIF model achieves IRs of 16.73% for machine-made sand and 23.54% for natural sand compared to the NI-MCIF model, and 32.1% for machine-made sand and 18.8% for natural sand compared to the ND-MCIF model. The NID-CC model achieves IR of 12.73% for machine-made sand and 12.93% for natural sand compared to the NI-CC model, and 17.77% for machine-made sand and 0.55% for natural sand compared to the ND-CC model. The results indicate that trimodal models based on simple concatenation fusion strategies exhibit significant variability across different types of sand materials. Notably, in the case of MC detection for natural sand, the performance of the trimodal model NID-CC is very close to that of the bimodal model ND-CC. This can be explained by the fact that models based on simple concatenation fusion fail to account for the heterogeneity and potential interactions between different modalities. Consequently, the detection results are more susceptible to interference from redundant features and noisy data, leading to no significant improvement in performance. It is worth noting that the ID-CC model outperforms the NID-CC, NI-CC, and ND-CC models significantly for both types of sand materials. This further supports the conclusion that simple concatenation-based fusion models fail to effectively extract critical features from different modalities, leading to either no noticeable performance improvement or even performance degradation as additional modalities are fused. In contrast, trimodal models based on the MCIF fusion strategy significantly outperform bimodal models under the same fusion strategy. Furthermore, bimodal models such as NI-MCIF and ND-MCIF, utilizing the MCIF fusion strategy, achieve substantially better performance than the trimodal model NID-CC. These results validate the effectiveness of the proposed fusion strategy in extracting key features from multiple modalities, considering the interactions among multimodal features, and mitigating the interference from noisy and redundant features

Table S5. Multimodal model performance statistics.

Model	Machine-made sand			Natural sand			Params	FLOPs
	R^2	RMSE	RPD	R^2	RMSE	RPD		
NID-MCIF	**0.983**	**0.458**	**7.9**	**0.984**	**0.705**	**7.931**	**39.633M**	**3.377G**
NID-CC	0.938	0.884	4.097	0.945	1.279	4.372	9.066M	3.315G
NID-DMF	0.963	0.706	5.128	0.975	0.876	6.386	8.161M	3.314G
NID-SAF	0.956	0.747	4.845	0.962	1.069	5.231	28.710M	3.355G
NI-CC	0.918	1.013	3.572	0.928	1.469	3.807	9.066M	3.316G
NI- MCIF	0.976	0.550	6.590	0.964	1.038	5.388	39.622M	3.375G
NI- DMF	0.947	0.817	4.429	0.954	1.170	4.782	8.161M	3.314G
NI- SAF	0.948	0.808	4.480	0.960	1.094	5.113	28.710M	3.355G
ND-CC	0.908	1.075	3.366	0.945	1.286	4.351	0.798M	5.678M
ND- MCIF	0.971	0.599	6.043	0.975	0.868	6.443	21.096M	44.815M
ND- DMF	0.956	0.747	4.847	0.967	0.995	5.625	0.131M	2.886M
ND- SAF	0.955	0.754	4.804	0.961	1.082	5.169	13.812M	30.254M
ID-CC	0.968	0.632	5.724	0.961	1.074	5.208	4.579M	0.829G

Section S4.3 Robustness analysis

Fig. S7 illustrates the variation in R^2 values for two types of sand datasets under different levels of image noise. As the image noise level increases, the proposed NID-MCIF model consistently maintains a high level of accuracy in

R^2, showing no significant downward trend. Furthermore, it demonstrates strong stability across multiple repeated experiments, and significantly outperforms both the NID-DMF, NID-SAF, and NID-CC models，indicating robust performance. For the NI-MCIF model, a gradual decline in R^2 is observed with increasing noise in the machine-made sand dataset, while a trend of initial increase followed by a decline is noted in the natural sand dataset. However, it consistently achieves high prediction accuracy, reflecting good robustness. These results confirm that the proposed MCIF-based fusion strategy renders the model insensitive to image noise, making it suitable for noisy image scenarios. In contrast, models employing a simple feature concatenation strategy, such as NID-CC and NI-CC, exhibit significant declines in R^2 as noise levels increase. This trend is particularly pronounced in the natural sand dataset, where performance shows considerable volatility, making these models less adaptive to noisy environments. The image-based model, IEfficientNet, also experiences a relatively obvious decrease in R^2 as noise levels rise, indicating weaker robustness. Notably, the ID-CC model demonstrates strong robustness, which can be attributed to the DC features providing more critical information and enhancing the model’s robustness.

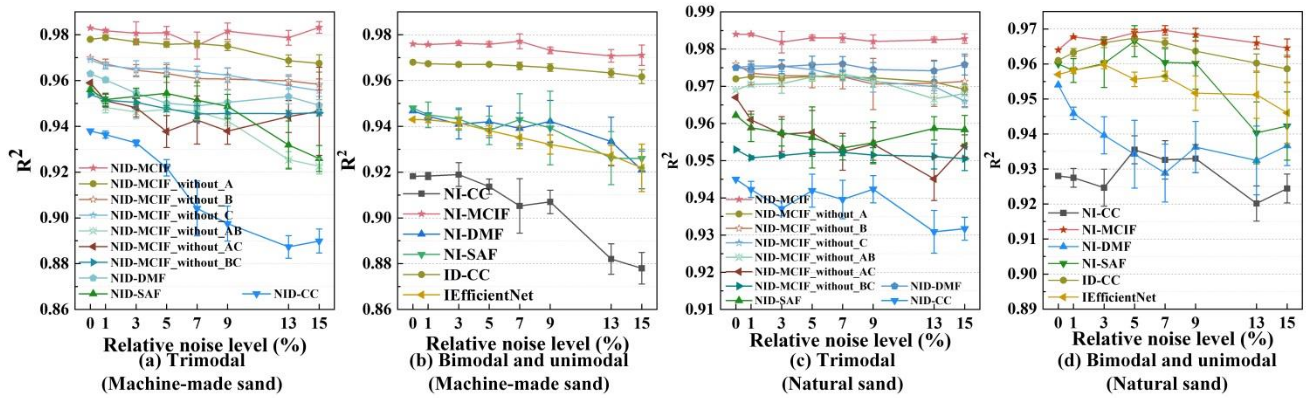

Fig. S7. Model performance under varying levels of image noise intensity.

Fig. S8 illustrates the variation in R^2 values for the two sand datasets under different levels of spectral noise. As spectral noise increases, the proposed NID-MCIF model consistently maintains a high level of R^2 accuracy without exhibiting any significant downward trend, demonstrating strong robustness. The NI-MCIF model shows strong robustness in the machine-made sand dataset. In the natural sand dataset, NI-MCIF exhibits strong noise resistance when the relative noise level is below 5%. However, as noise increases further, its noise resistance declines. The ND-MCIF model displays no significant downward trend across different noise levels for both sand types, indicating strong robustness. Nevertheless, in the natural sand dataset, certain fluctuations are observed in repeated experiments under medium to high noise levels. In contrast, models based on simple concatenation fusion strategies, such as NID-CC, NI-CC, and ND-CC, exhibit pronounced downward trends and considerable fluctuations, making them less adaptable to noisy environments. Unimodal models generally show a clear downward trend and struggle to adapt to noise. Additionally, under different modal combinations (tri-modal and bi-modal), the MCIF model demonstrates superior noise robustness compared to the DMF and SAF models. These findings indicate that the proposed MCIF fusion strategy achieves the strongest robustness with trimodal data and retains strong robustness with bimodal data under low to medium spectral noise levels. However, in unimodal settings, the effectiveness of the MCIF model is limited due to the constrained number of features.

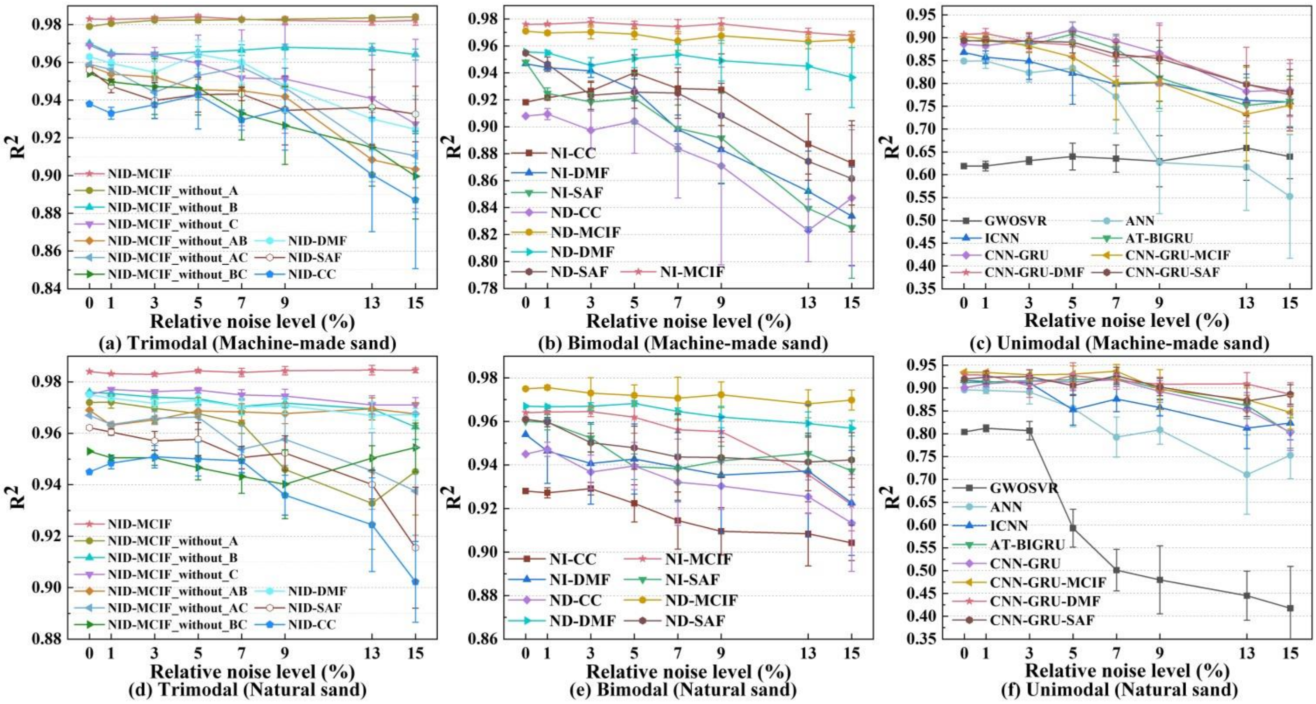

Fig. S8. Model performance under varying levels of NIR noise intensity.

Fig. S9 illustrates the variation in R^2 values for the two sand datasets under different levels of DC noise. When the relative DC noise level is less than or equal to 5%, the proposed NID-MCIF model exhibits a gradual downward trend in R^2 while maintaining higher accuracy compared to other models. However, as noise levels continue to increase, the model's noise resistance declines. The ND-MCIF model follows a similar trend to the NID-MCIF model. Models such as NID-DMF, NID-SAF, NID-CC, ND-DMF, ND-SAF, ND-CC, and ID-CC demonstrate relatively strong stability. In contrast, unimodal models show a pronounced downward trend, making them less adaptable to noisy environments. These results indicate that DC features play a crucial role in the MCIF fusion strategy. When the relative noise level is within a low to medium range (less than or equal to 7%), the proposed NID-MCIF model can maintain an R^2 value above 97%, demonstrating strong noise resistance. Conversely, models based on simple concatenation fusion strategies are less affected by noise due to the lack of consideration for interactions between different modalities. This limitation contributes to their relatively low accuracy under low noise levels.

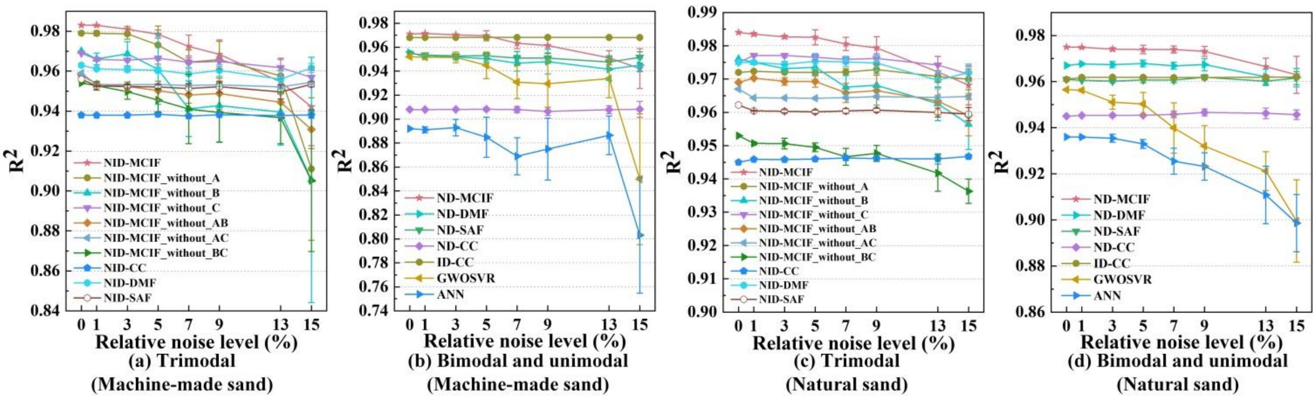

Fig. S9. Model performance under varying levels of DC noise intensity.