17th World Conference on Earthquake Engineering, 17WCEE

Sendai, Japan - September 13th to 18th 2020

Study on damage estimation by machine learning of relationship between building response and damage

A. Aoi[1], H. Tsunekawa[2], M. Yoshizawa[3], A. Kanbayashi[4], G. Nyamkhuu [5], S. Tano [6]

[1] *Associate Chief Researcher, Takenaka Corporation, aoi.atsushi@takenaka.co.jp*
[2] *Group Leader, Takenaka Corporation, tsunekawa.hiroshi@takenaka.co.jp*
[3] *Group Leader, Takenaka Corporation, yoshizawa.mutsuhiro@takenaka.co.jp*
[4] *Chief Researcher, Takenaka Corporation, kanbayashi.atsushi@takenaka.co.jp*
[5] *Graduate Student, The University of Electro-Communications, g1830033@edu.cc.uec.ac.jp*
[6] *Prof., The University of Electro-Communications, tano@is.uec.ac.jp*

Abstract

Post-earthquake use of buildings is required at hospital facilities that responds to disaster medical care and in measures for stranded persons, and so on. In the Great East Japan Earthquake, stranded persons were a major issue, and the building is required to promptly confirm the safety of the facility after the earthquake in order to accept stranded persons, and the need for structural health monitoring system is increasing in recent years. Many of these systems install accelerometers on each floor or every few floors in the building, estimate the building deformation by acceleration integration or mode analysis, and estimate the damage of each floor from story drift angle.

The authors speculated that the degree of building damage could be estimated directly by training with machine learning in advance the relationship between the acceleration data measured by accelerometers installed in the building and the degree of building damage. However, in order to create a learning model, a huge amount of training data including the results of damage is necessary, and observation records alone are insufficient. Therefore, we considered to create a learning model by conducting earthquake response analysis many times and using the analysis results and observation records as training data.

In this study, as the first step, we created a learning model using only the analysis results. Since the target building is preferably that with the observation records, to create training data, 3D model for seismic response analysis was created referring to the specimen used in the previous full-scale shake table experiment conducted in E-Defense for a steel 18-story building, and seismic response analysis with multiple seismic waves have been conducted.

Machine learning model is trained using the Fourier amplitude spectrum of each floor response acceleration obtained from the analysis results as input data, and the damage level of each floor as output data. The damage level of each floor is categorized into 3 levels based on the rate of structural member hinge occurrence.

As a result of machine learning with neural networks and Random Forest, it was found that the damage level of each floor can be estimated with accuracy of more than 80%. However, the estimated accuracy of the damage pattern that was infrequent in the training data was low. When the feature importance was evaluated with Random Forest, the importance of the Fourier amplitude spectrum corresponding to the position where the seismic response is not amplified tended to be high.

Keywords: building damage, earthquake, machine learning, neural network, Random Forest

1. Introduction

Post-earthquake use of buildings is required at hospital facilities that responds to disaster medical care and in measures for stranded persons, and so on. In the Great East Japan Earthquake, stranded persons were a major issue, and the building is required to promptly confirm the safety of the facility after the earthquake in order to accept stranded persons. Thus, in 2013, Ordinance for Measures Concerning Stranded Persons was enacted by the Tokyo Metropolitan government. According to the final report by the Council for Measures on Stranded People Concerning the Capital Region Earthquake [1], it is required to judge within three hours of whether or not to stay in the building, and the need for structural health monitoring system is increasing in recent years. Many of these systems install accelerometers on each floor or every few floors in the building, estimate the building deformation by acceleration integration or mode analysis, and estimate the damage of each story from story drift angle. These types are also being introduced into real buildings as systems with some degree of reliability.

The authors speculated that the degree of building damage could be estimated directly by training with machine learning in advance the relationship between the acceleration data measured by accelerometers installed in the building and the degree of building damage. With this method, not only each story but also each member can be estimated. However, in order to create a learning model, a huge amount of training data including the results of damage is necessary, and observation records alone are insufficient. Therefore, we considered to create a learning model by conducting seismic response analysis many times and using the analysis results and observation records as training data. This paper represents the result of creation a learning model to estimate the degree of damage each story using only the analysis results as the first step.

2. Create Training Data

2.1 Target Building

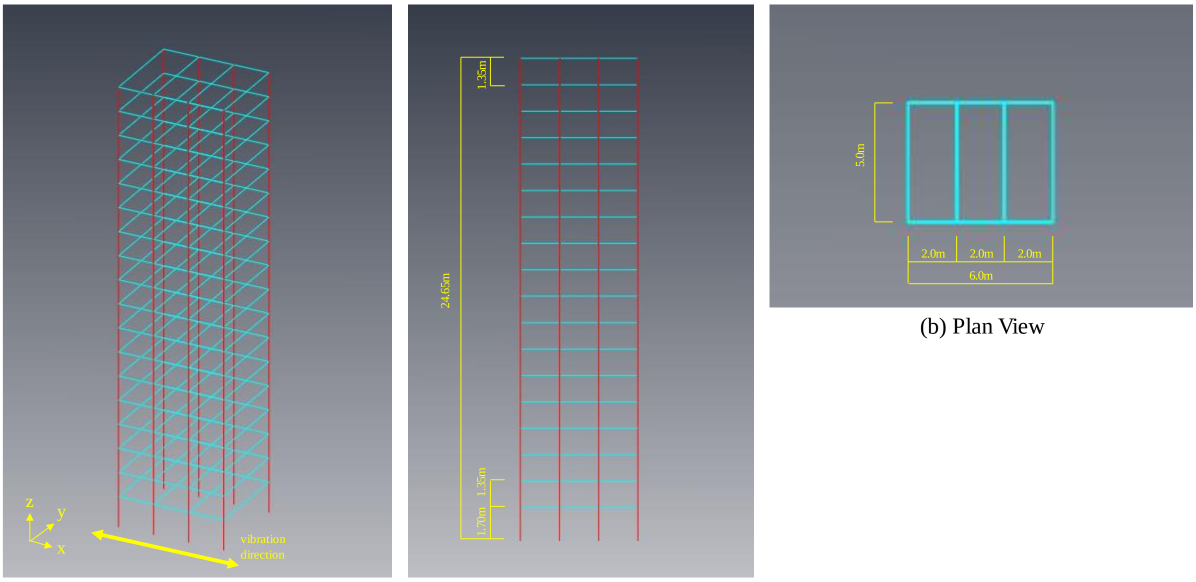
Since the target building is preferably that with the observation records, to create training data, 3D model for seismic response analysis was created referring to the specimen used in the previous full-scale shake table experiment conducted in E-Defense for a steel 18-story building [2]. Table 1 shows the specifications of the created analysis model, Table 2 shows the eigenvalue analysis results, Fig.1 shows the analysis model diagram, and Fig.2 shows story shear force－story drift relationship of pushover analysis.

Table 1 – Specification of the analysis model

Item	Specification
Structure	18 story steel frame structure
Gross weight	4179kN
height	25.35m
Plan Size	6.0m×5.0m(2×3)
scale	1/3

Table 2 – Eigenvalue analysis results

mode order (X-direction)	Period (sec)	Frequency (Hz)	Effective mass ratio			
			X	Y	Z	θZ
1	1.14	0.875	0.770	0.000	0.000	0.191
2	0.37	2.681	0.138	0.000	0.000	0.034
3	0.20	4.897	0.038	0.000	0.000	0.009



(a) Over View (c) Elevation (b) Plan View

Fig. 1 – 3D model for seismic response analysis

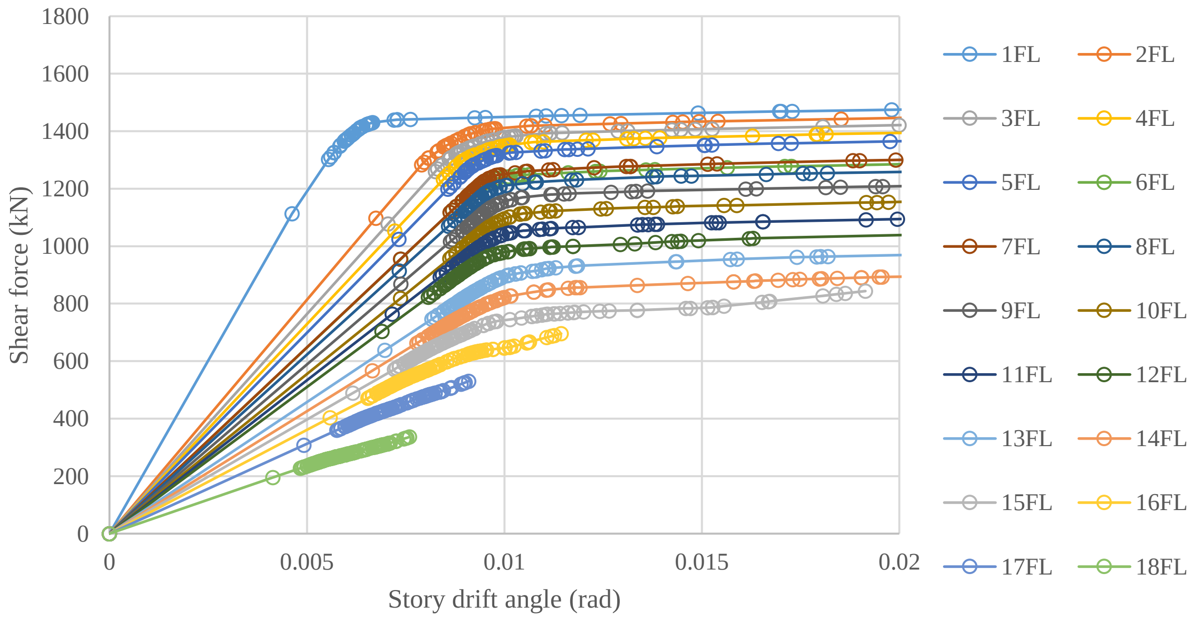


Fig. 2 – Story shear force－story drift relationship of pushover analysis (X-direction)

2.2 Seismic Response Analyses

Training data is created by seismic response analysis with multiple seismic waves on the above analysis model. The ground motion used in the analysis is observed at K-NET Naruko station (MYG005), where the ground surface amplification peaked around the first mode period of the analysis model. 140 waves were selected from observation records with JMA seismic intensity of over 2.0 observed the period from January 1, 1996 to September 4, 2017. The analysis is performed in one direction (X-direction), and the maximum direction of composite vector of the horizontal two-direction of 140 waves is used. Fig.3 shows velocity response spectrums of all observation records.

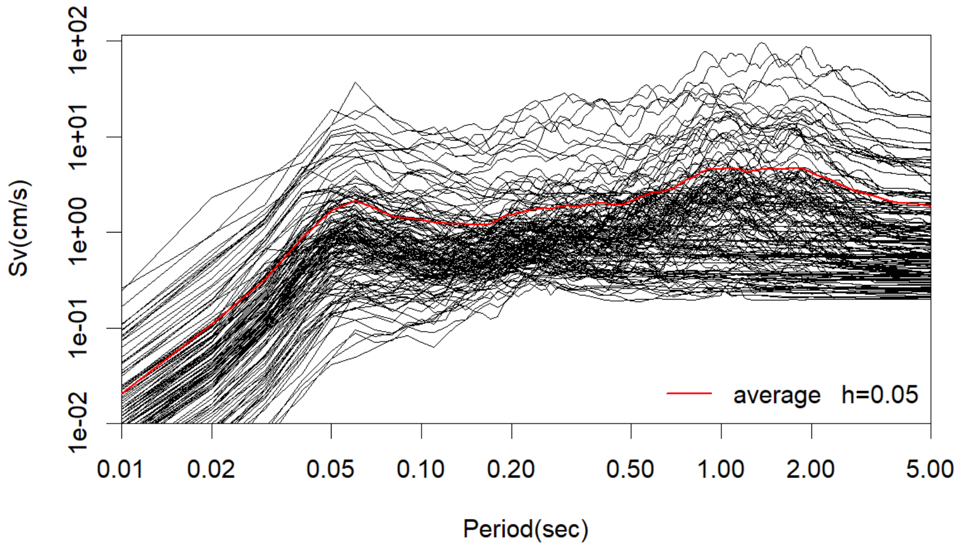


Fig. 3 – Velocity response spectrums of 140 waves

The above observation record was standardized so that the maximum velocity was 115 cm/s, and a waveform was prepared for each 100 waves in 1% increments from an amplitude level of 1% to 100% (14,000 waves in total). In the experiment, since the test specimen was reduced to 1/3 scale compared to the one that assumed an 18-story building, the time-step of the ground motion was also scaled to 1/√3. A case where the response becomes too large is regarded as an error case, and a result where the maximum story drift angle or the member ductility factor is too large is excluded from the training data. The case where the story drift angle exceeds 1/20 or the member plasticity exceeds 50 are excluded, and 12,770 cases remained. Fig.4 shows the maximum story drift angle distribution for 12,770 cases.

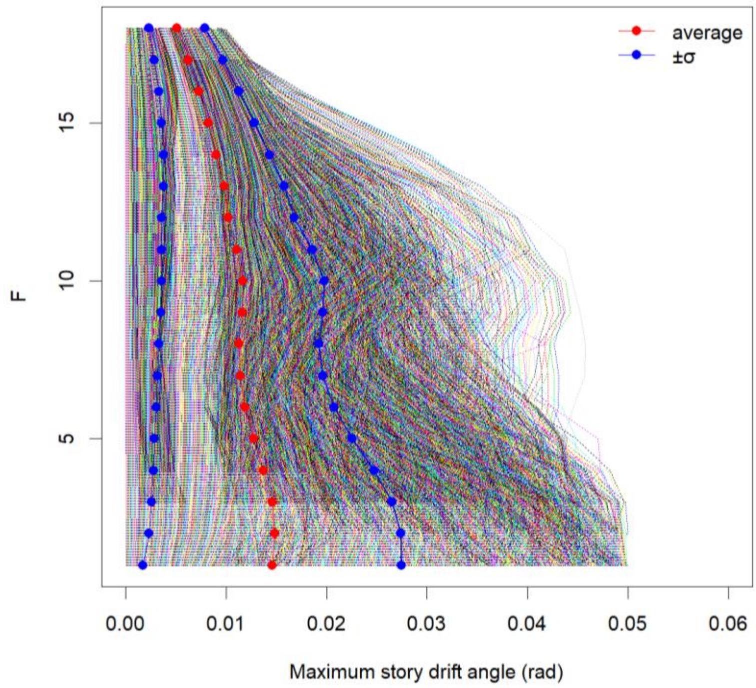
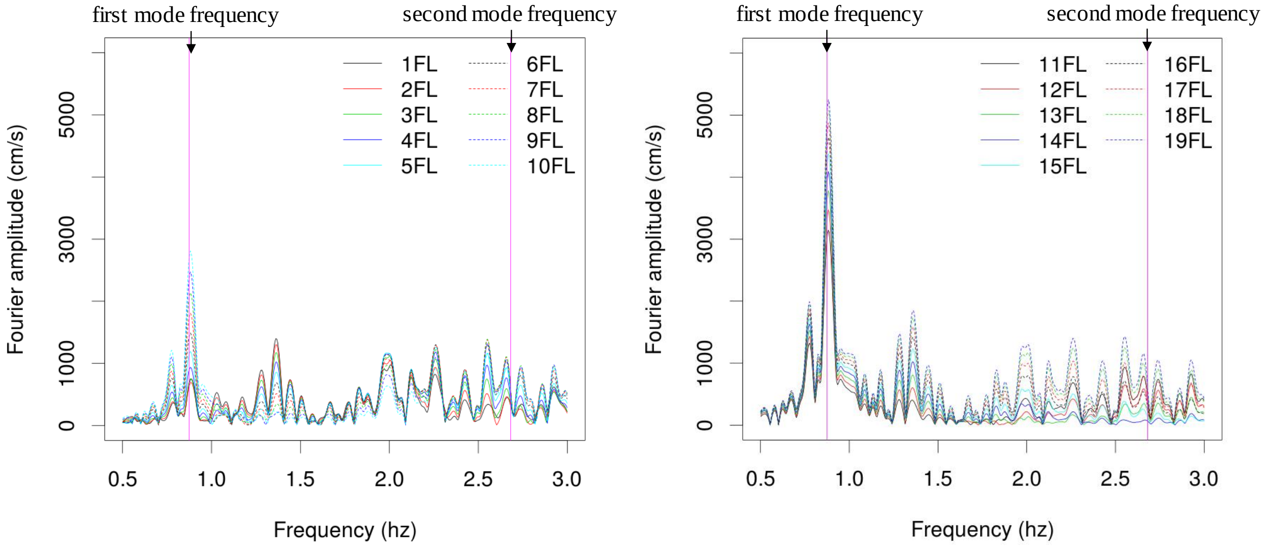


Fig. 4 – Maximum story drift angle distribution for 12,770 cases

2.3 Input / Output Data

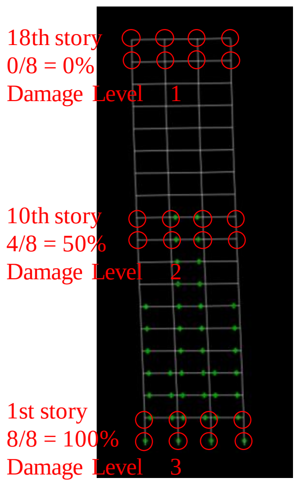
As the input data, the Fourier amplitude of the 1st to 19th floor acceleration data is adopted, considering that the response in the frequency domain is correlated with the damage to the building. The target frequency band was 0.5 to 3.0 Hz, 19 × 474 (0.00527 Hz steps) matrix data was used as input data. Fig.5 shows an example of Fourier amplitude spectrum when the amplitude is 100% in the seismic wave No.1.

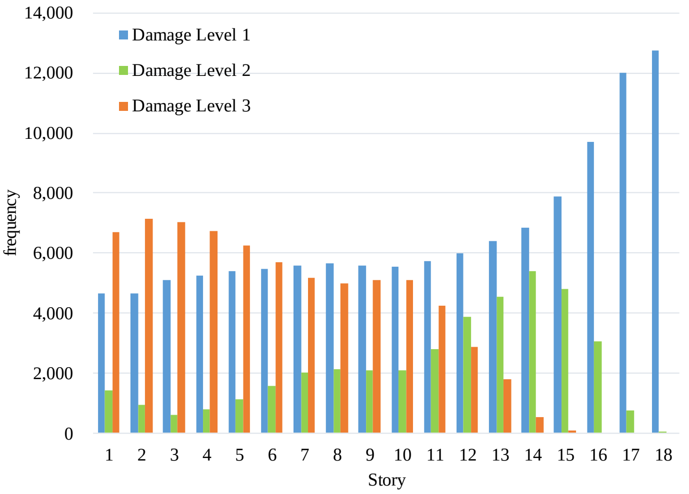


(a) 1st to 10th floor (b) 11th to 19th floor

Fig. 5 – Fourier amplitude spectrum (seismic wave No.1, 100% amplitude)

The output data was defined as the damage level of each story, and defined by the occurrence ratio of plastic hinges at the nodes on the upper and lower floors of each story. If a plastic hinge occurs (the member ductility factor exceeds 1.0) at any of the member ends in contact with the node, the plastic hinge is counted as having occurred at the node. The damage level was defined as three levels of plastic hinge occurrence ratio of 0%, 0 to 100%, and 100%. Figure 6 shows an example of determining the damage level and shows the frequency of each damage level on each story in 12,770 cases. It can be seen that the ratio of level 2 is lower than that of levels 1 and 3, and the damage level is tend to be higher in lower stories than in higher stories.





Story	Damage Level		
	1	2	3
18	12,757	13	0
17	12,013	757	0
16	9,712	3,058	0
15	7,895	4,799	76
14	6,865	5,385	520
13	6,400	4,560	1,810
12	5,987	3,896	2,887
11	5,734	2,788	4,248
10	5,562	2,089	5,119
9	5,577	2,098	5,095
8	5,655	2,118	4,997
7	5,567	2,015	5,188
6	5,492	1,578	5,700
5	5,385	1,123	6,262
4	5,240	798	6,732
3	5,113	615	7,042
2	4,675	954	7,141
1	4,675	1,411	6,684

(a) example of determining the damage level (b) frequency of each damage level on each story in 12,770 cases

Fig. 6 – The damage level of each story

3. Machine Learning

3.1 Learning Model

Learning is performed using three models, Multi-Layer Perceptron (MLP), Convolutional Neural Network (CNN), and Random Forest (RF). The learning model is set individually for each story damage level, and a total of 18 learning models are constructed. Fig.7 shows the network architecture of MLP and CNN. The number of decision trees created in RF are 400, and the number of features used to create each decision tree is determined by grid search. The training data is divided into the training set and the test set. Since the test set is need to be unknown seismic waveforms for the training set, No.1-35 seismic waves cases are prepared for the test set (3,154 case), and the other 105 waves cases are used as the training set (9,616 case), and training and evaluation are performed with the dataset.

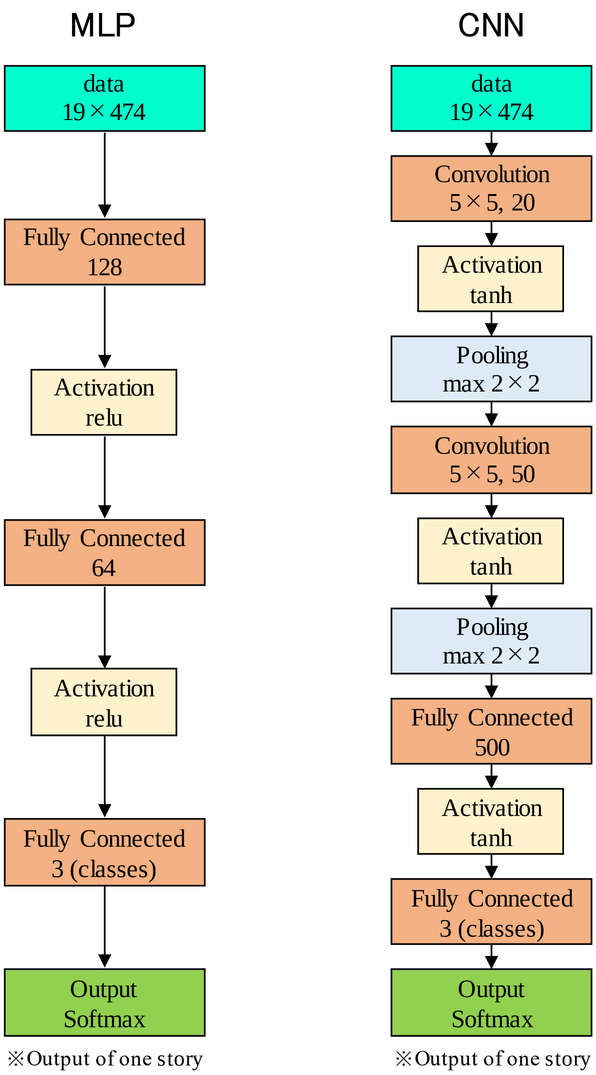


Fig. 7 – The neural network architecture

3.2 Learning Result

The accuracy is evaluated for each learning model. Table 3 shows the result of comparing the predicted value and the measured value of each floor damage level of each learning model, and Fig.8 shows the result of comparing the accuracy of each learning model.

Table 3 – Comparison table of measured and predicted damage level

	MLP				CNN				RF			
1F	predicted	Accuracy =		0.8827	predicted	Accuracy =		0.8786	predicted	Accuracy =		0.9008
	measured	1	2	3	measured	1	2	3	measured	1	2	3
	1	1027	72	5	1	1056	42	6	1	1040	64	0
	2	85	159	97	2	119	100	122	2	70	159	112
	3	43	68	1598	3	31	63	1615	3	2	65	1642
2F	predicted	Accuracy =		0.8884	predicted	Accuracy =		0.8859	predicted	Accuracy =		0.9039
	measured	1	2	3	measured	1	2	3	measured	1	2	3
	1	1030	63	11	1	1074	0	30	1	1054	23	27
	2	54	95	94	2	101	1	141	2	63	47	133
	3	79	51	1677	3	86	2	1719	3	20	37	1750
3F	predicted	Accuracy =		0.9017	predicted	Accuracy =		0.9036	predicted	Accuracy =		0.9166
	measured	1	2	3	measured	1	2	3	measured	1	2	3
	1	1152	30	39	1	1173	0	48	1	1166	2	53
	2	49	31	75	2	69	0	86	2	57	6	92
	3	91	26	1661	3	101	0	1677	3	53	6	1719
4F	predicted	Accuracy =		0.8909	predicted	Accuracy =		0.8944	predicted	Accuracy =		0.9115
	measured	1	2	3	measured	1	2	3	measured	1	2	3
	1	1173	56	27	1	1207	0	49	1	1204	7	45
	2	56	44	91	2	85	0	106	2	58	17	116
	3	70	44	1593	3	93	0	1614	3	39	14	1654
5F	predicted	Accuracy =		0.8725	predicted	Accuracy =		0.8843	predicted	Accuracy =		0.9093
	measured	1	2	3	measured	1	2	3	measured	1	2	3
	1	1202	79	18	1	1248	15	36	1	1235	56	8
	2	68	75	107	2	100	20	130	2	64	104	82
	3	53	77	1475	3	69	15	1521	3	37	39	1529
6F	predicted	Accuracy =		0.8703	predicted	Accuracy =		0.8627	predicted	Accuracy =		0.9084
	measured	1	2	3	measured	1	2	3	measured	1	2	3
	1	1242	91	15	1	1262	76	10	1	1257	90	1
	2	69	187	100	2	91	95	170	2	51	214	91
	3	33	101	1316	3	34	52	1364	3	18	38	1394
7F	predicted	Accuracy =		0.8618	predicted	Accuracy =		0.8488	predicted	Accuracy =		0.8982
	measured	1	2	3	measured	1	2	3	measured	1	2	3
	1	1265	102	2	1	1264	104	1	1	1260	109	0
	2	88	305	83	2	95	199	182	2	58	326	92
	3	27	134	1148	3	22	73	1214	3	14	48	1247
8F	predicted	Accuracy =		0.8507	predicted	Accuracy =		0.8450	predicted	Accuracy =		0.8827
	measured	1	2	3	measured	1	2	3	measured	1	2	3
	1	1283	109	1	1	1284	108	1	1	1266	127	0
	2	103	332	82	2	92	243	182	2	62	340	115
	3	26	150	1068	3	20	86	1138	3	9	57	1178
9F	predicted	Accuracy =		0.8548	predicted	Accuracy =		0.8519	predicted	Accuracy =		0.8906
	measured	1	2	3	measured	1	2	3	measured	1	2	3
	1	1264	110	0	1	1274	97	3	1	1263	110	1
	2	97	341	88	2	93	262	171	2	61	359	106
	3	24	139	1091	3	20	83	1151	3	3	64	1187
10F	predicted	Accuracy =		0.8503	predicted	Accuracy =		0.8649	predicted	Accuracy =		0.9033
	measured	1	2	3	measured	1	2	3	measured	1	2	3
	1	1257	115	0	1	1267	105	0	1	1259	113	0
	2	90	327	90	2	82	302	123	2	44	382	81
	3	35	142	1098	3	29	87	1159	3	12	55	1208
11F	predicted	Accuracy =		0.8237	predicted	Accuracy =		0.8510	predicted	Accuracy =		0.8773
	measured	1	2	3	measured	1	2	3	measured	1	2	3
	1	1284	125	0	1	1294	115	0	1	1284	125	0
	2	121	493	115	2	104	511	114	2	54	569	106
	3	45	150	821	3	24	113	879	3	0	102	914
12F	predicted	Accuracy =		0.8196	predicted	Accuracy =		0.8456	predicted	Accuracy =		0.8748
	measured	1	2	3	measured	1	2	3	measured	1	2	3
	1	1358	110	0	1	1350	118	0	1	1354	114	0
	2	150	793	115	2	115	869	74	2	67	903	88
	3	36	158	434	3	14	166	448	3	0	126	502
13F	predicted	Accuracy =		0.8386	predicted	Accuracy =		0.8488	predicted	Accuracy =		0.8808
	measured	1	2	3	measured	1	2	3	measured	1	2	3
	1	1491	70	0	1	1449	112	0	1	1450	111	0
	2	210	993	40	2	138	1078	27	2	76	1131	36
	3	11	178	161	3	0	200	150	3	0	153	197
14F	predicted	Accuracy =		0.8754	predicted	Accuracy =		0.8843	predicted	Accuracy =		0.9109
	measured	1	2	3	measured	1	2	3	measured	1	2	3
	1	1614	73	0	1	1568	119	0	1	1581	106	0
	2	178	1127	52	2	141	1213	3	2	66	1291	0
	3	0	90	20	3	0	102	8	3	0	109	1
15F	predicted	Accuracy =		0.9081	predicted	Accuracy =		0.9242	predicted	Accuracy =		0.9464
	measured	1	2	3	measured	1	2	3	measured	1	2	3
	1	1859	100	0	1	1864	95	0	1	1870	89	0
	2	184	1005	0	2	138	1051	0	2	74	1115	0
	3	0	6	0	3	0	6	0	3	0	6	0
16F	predicted	Accuracy =		0.9109	predicted	Accuracy =		0.9318	predicted	Accuracy =		0.9483
	measured	1	2	3	measured	1	2	3	measured	1	2	3
	1	2386	73	0	1	2427	32	0	1	2384	75	0
	2	208	487	0	2	183	512	0	2	88	607	0
	3	0	0	0	3	0	0	0	3	0	0	0
17F	predicted	Accuracy =		0.9540	predicted	Accuracy =		0.9597	predicted	Accuracy =		0.9569
	measured	1	2	3	measured	1	2	3	measured	1	2	3
	1	2968	41	0	1	3009	0	0	1	2998	11	0
	2	104	41	0	2	127	18	0	2	125	20	0
	3	0	0	0	3	0	0	0	3	0	0	0
18F	predicted	Accuracy =		1.0000	predicted	Accuracy =		1.0000	predicted	Accuracy =		1.0000
	measured	1	2	3	measured	1	2	3	measured	1	2	3
	1	3154	0	0	1	3154	0	0	1	3154	0	0
	2	0	0	0	2	0	0	0	2	0	0	0
	3	0	0	0	3	0	0	0	3	0	0	0

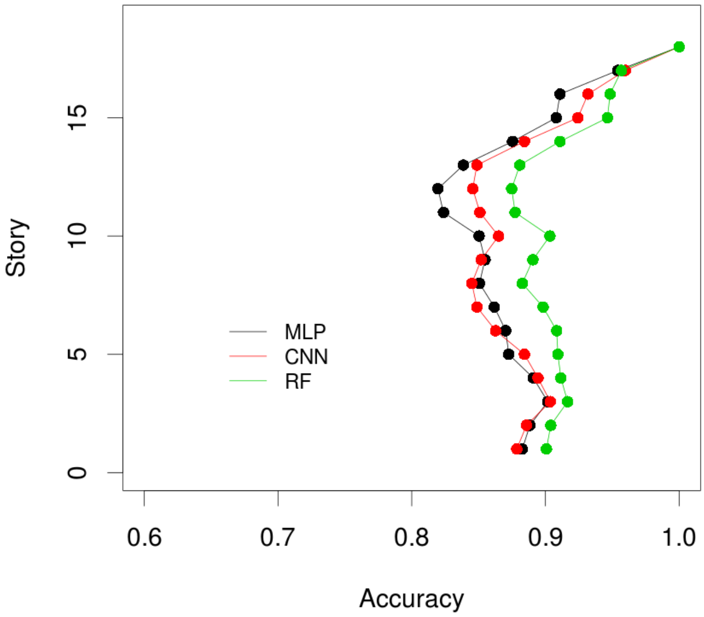


Fig. 8 – Comparing the accuracy of each learning model

There is tendency that the accuracy of 6th to 13th stories is lower than that of the other stories in any of the learning model results. The damage level 2 is considered to have lower accuracy due to the small number of data itself, and the middle stories has a higher damage level 2 ratio than other stories, so the accuracy is likely to be lower. The higher stories (14th story or higher) have a higher accuracy, but this is a story where damage is unlikely to occur, so damage level 3 ratio is small and the ratio of damage level 1 is very large, so accuracy is high. As a comparison between the models, there is little difference in the accuracy between the two types of neural networks. On the other hand, Random Forest is more accurate on almost all floors than the other two models.

3.3 Feature Importance

In the random forest, the feature importance can be evaluated. Fig.9 shows, for the RF model of each story, the input data of the top 10 importance levels among the input data of the 19 × 474 Fourier amplitude. The importance is based on the degree of decrease in the Gini coefficient. Except for the prediction models from the 16th story to the 18th story, almost stories tend to have a higher importance of the Fourier amplitude in the band from first mode frequency (0.875 Hz) to about 1.00 Hz in the 1st to 4th floors. This suggests that most stories may be able to predict damage using only lower floor acceleration data without for all floors. On the other hand, on the 16th story model, the importance of the Fourier amplitude of 1.0 to 1.3 Hz on the 2nd to 7th floors is high, and on the 17th story model, the importance of the Fourier amplitude of 1.0 to 1.7 Hz on the 3rd to 10th floors is high. There is a tendency that the higher the target story, the higher the floor level and the frequency of the Fourier amplitude having the high importance. As for the 18th floor, since the damage level is 1 in all cases, indicating that the importance of the feature value is meaningless.

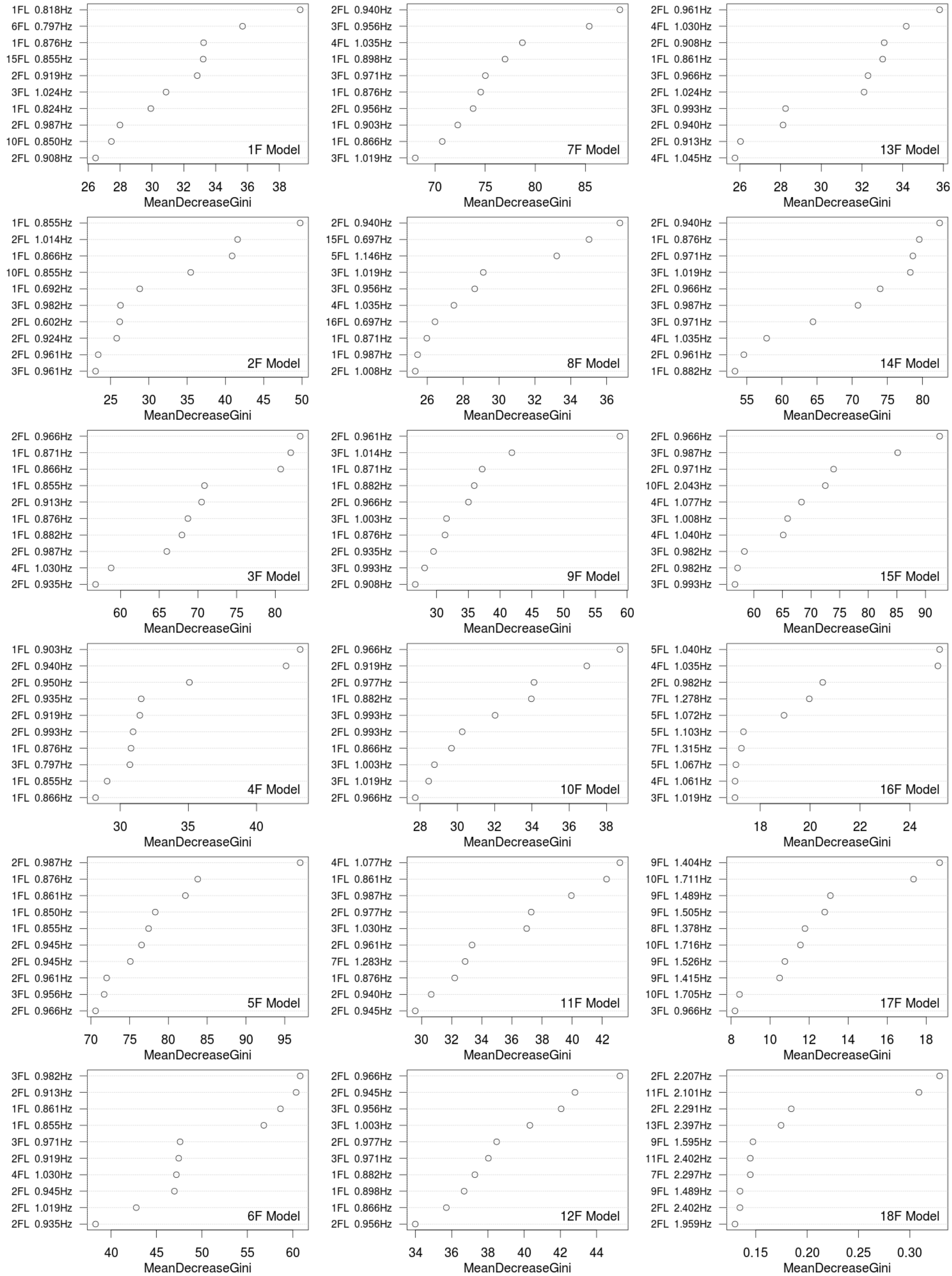


Fig. 9 – The feature having the top 10 importance levels

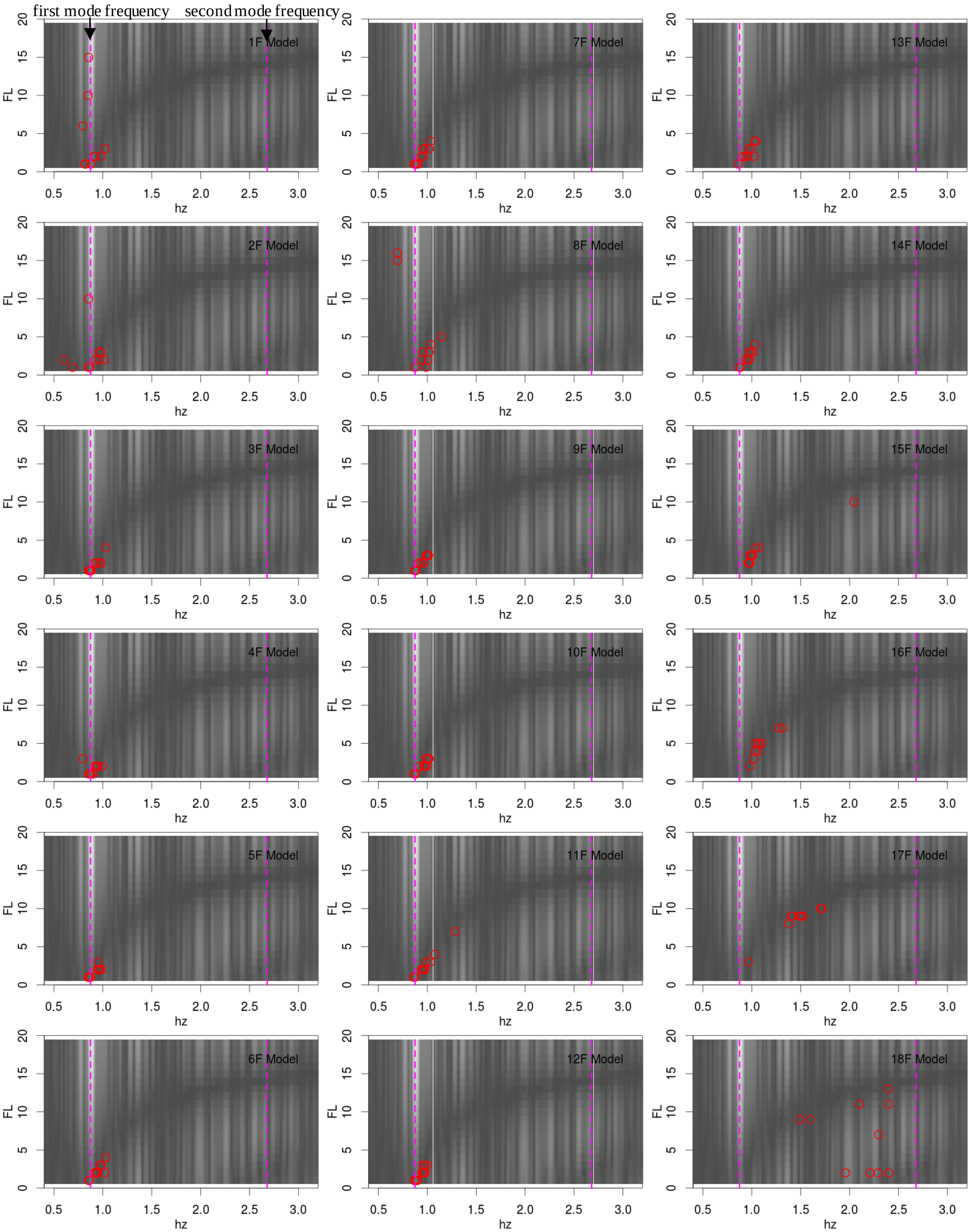


Fig. 10 – The Fourier amplitude spectrum of each floor overlaid with feature of the top 10 importance levels (The whiter the higher the Fourier amplitude, the blacker the smaller)

In order to confirm the relationship between the Fourier amplitude spectrum of each floor and the importance of the feature, Fig.10 shows a color map of the Fourier amplitude spectrum of each floor overlaid with feature of the top 10 importance levels (when the amplitude is 100% in the seismic wave No.1). It can be seen that, in addition to the tendency described in Fig.9, the position of the feature with high importance overlaps the band with a small Fourier amplitude. The Fourier amplitude at a frequency close to the first mode frequency of the ground motion is important, and the Fourier amplitude in an unresponsive frequency domain at above floors is also important. It is presumed that this learning model predicts damage based on the ground motion power at the first mode frequency of building and the acceleration amplitude in the frequency band not affected by the building response on the upper floor.

4. Conclusions

We speculated that it is possible to estimates the degree of building damage from acceleration data observed in buildings by training with machine learning the relationship between building response acceleration and building damage. Therefore, as the first step, machine learning was performed using the seismic response analysis results as training data. As a result, it was found that the analysis made it possible to estimate the damage of each story of the building from the acceleration data with accuracy of 80% or more by machine learning using a neural network or random forest. As a result of analyzing the importance of the feature, it was found that the ground motion power at the first mode frequency of the building and the acceleration amplitude in the frequency band not affected by the response of the building on the upper floor are important. The future task is how to solve the problem that the accuracy of the damage level with low occurrence frequency decreases because the occurrence frequency of each damage level greatly differs depending on the story. In addition, since this study is based on the analysis results, it will be necessary to verify in the future whether it is possible to make estimations using actual observation data.

5. Acknowledgements

Strong motion records were provided by NIED. For the building data, we used the experimental data of "Quantification of collapse margin of high-rise steel-framed buildings" from NIED "ASEBI".

6. References

[1] Cabinet Office, Government of Japan (2012): the final report by the Council for Measures on Stranded People Concerning the Capital Region Earthquake.

[2] NIED "ASEBI", https://www.edgrid.jp, "Quantification of collapse margin of high-rise steel-framed buildings"

