Harmonizing Data: Multilingual, Multistep Deep Learning Approach for Classifying Audio Content: Distinguishing Among Songs, Podcasts (Talk), and Advertisement

Details Classification Result

1.1 Binary Classification Models Result: Song Vs Podcast

Table 1: Song Vs Podcast Using Method 1(MFCC Feature) for English Data

Model	F1	Confusion Matrix	Classification Report
CNN	.96	<table><tr><td></td><td>Podcast</td><td>Song</td></tr><tr><td>Podcast</td><td>9722</td><td>85</td></tr><tr><td>Song</td><td>278</td><td>4892</td></tr></table>	<table><tr><td>Class</td><td>Precision</td><td>Recall</td><td>F1-Score</td><td>Support</td></tr><tr><td>PodCast</td><td>.99</td><td>.97</td><td>.98</td><td>10000</td></tr><tr><td>Song</td><td>.95</td><td>.98</td><td>.96</td><td>4977</td></tr></table>
VGG Net	.94	<table><tr><td></td><td>Podcast</td><td>Song</td></tr><tr><td>Podcast</td><td>9678</td><td>253</td></tr><tr><td>Song</td><td>322</td><td>4724</td></tr></table>	<table><tr><td>Class</td><td>Precision</td><td>Recall</td><td>F1-Score</td><td>Support</td></tr><tr><td>PodCast</td><td>.97</td><td>.97</td><td>.97</td><td>10000</td></tr><tr><td>Song</td><td>.94</td><td>.95</td><td>.94</td><td>4977</td></tr></table>
Densenet	.38	<table><tr><td></td><td>Podcast</td><td>Song</td></tr><tr><td>Podcast</td><td>9917</td><td>3785</td></tr><tr><td>Song</td><td>83</td><td>1192</td></tr></table>	<table><tr><td>Class</td><td>Precision</td><td>Recall</td><td>F1-Score</td><td>Support</td></tr><tr><td>PodCast</td><td>.72</td><td>.99</td><td>.84</td><td>10000</td></tr><tr><td>Song</td><td>.93</td><td>.24</td><td>.38</td><td>4977</td></tr></table>
MobileNet	.81	<table><tr><td></td><td>Podcast</td><td>Song</td></tr><tr><td>Podcast</td><td>8056</td><td>284</td></tr><tr><td>Song</td><td>1944</td><td>4693</td></tr></table>	<table><tr><td>Class</td><td>Precision</td><td>Recall</td><td>F1-Score</td><td>Support</td></tr><tr><td>PodCast</td><td>.97</td><td>.81</td><td>.88</td><td>10000</td></tr><tr><td>Song</td><td>.71</td><td>.94</td><td>.81</td><td>4977</td></tr></table>

Table 2: Song Vs Podcast Using Method 2 (Mel-Spectrogram Applied decibel scaling to data) for English Data

Model	F1	Confusion Matrix	Classification Report
CNN	.97	<table><tr><td></td><td>Podcast</td><td>Song</td></tr><tr><td>Podcast</td><td>9771</td><td>45</td></tr><tr><td>Song</td><td>229</td><td>4933</td></tr></table>	<table><tr><td>Class</td><td>Precision</td><td>Recall</td><td>F1-Score</td><td>Support</td></tr><tr><td>PodCast</td><td>1</td><td>.98</td><td>.99</td><td>10000</td></tr><tr><td>Song</td><td>.96</td><td>.99</td><td>.97</td><td>4978</td></tr></table>
VGG Net	.98	<table><tr><td></td><td>Podcast</td><td>Song</td></tr><tr><td>Podcast</td><td>9835</td><td>28</td></tr><tr><td>Song</td><td>165</td><td>4950</td></tr></table>	<table><tr><td>Class</td><td>Precision</td><td>Recall</td><td>F1-Score</td><td>Support</td></tr><tr><td>PodCast</td><td>1</td><td>.98</td><td>.99</td><td>10000</td></tr><tr><td>Song</td><td>.97</td><td>.99</td><td>.98</td><td>4978</td></tr></table>
Densenet	.95	<table><tr><td></td><td>Podcast</td><td>Song</td></tr><tr><td>Podcast</td><td>9808</td><td>255</td></tr></table>	<table><tr><td>Class</td><td>Precision</td><td>Recall</td><td>F1-Score</td><td>Support</td></tr><tr><td>PodCast</td><td>.97</td><td>.98</td><td>.98</td><td>10000</td></tr></table>

Model	F1	Confusion Matrix			Classification Report				
			Podcast	Song	Song	.96	.95	.95	4978
		Song	192	4723					
MobileNet	.96		Podcast	Song	Class	Precision	Recall	F1-Score	Support
		Podcast	9735	60	PodCast	.99	.97	.98	10000
		Song	265	4918	Song	.95	.99	.97	4978

Table 3: Song Vs Podcast Using Method 1(MFCC Feature) for Hindi Data

Model	F1	Confusion Matrix			Classification Report				
CNN	.89		Podcast	Song	Class	Precision	Recall	F1-Score	Support
		Podcast	8924	85	PodCast	.99	.89	.94	10000
		Song	1076	4915	Song	.82	.98	.89	5000
VGG Net	.94		Podcast	Song	Class	Precision	Recall	F1-Score	Support
		Podcast	9694	266	PodCast	.97	.97	.97	10000
		Song	306	4734	Song	.94	.94	.95	5000
Densenet	.28		Podcast	Song	Class	Precision	Recall	F1-Score	Support
		Podcast	9993	4196	PodCast	.70	1	.83	10000
		Song	7	804	Song	.99	.16	.28	5000
MobileNet	.72		Podcast	Song	Class	Precision	Recall	F1-Score	Support
		Podcast	9939	2127	PodCast	.82	.99	.90	10000
		Song	61	2873	Song	.98	.57	.72	5000

Table 4: Song Vs Podcast Using Method 2 (Mel-Spectrogram Applied decibel scaling to data) for Hindi Data

Model	F1	Confusion Matrix			Classification Report				
CNN	.98		Podcast	Song	Class	Precision	Recall	F1-Score	Support
		Podcast	9913	61	PodCast	.99	.99	.99	10000
		Song	87	4939	Song	.98	.99	.99	5000
VGG Net	.99		Podcast	Song	Class	Precision	Recall	F1-Score	Support
		Podcast	9961	76	PodCast	.99	1	.99	10000
		Song	39	4924	Song	.99	.98	.98	5000

Model	F1	Confusion Matrix				Classification Report				
Densenet	.96		Podcast	Song		Class	Precision	Recall	F1-Score	Support
		Podcast	9905	244		PodCast	.98	.99	.98	10000
		Song	95	4756		Song	.98	.95	.97	5000
MobileNet	.94		Podcast	Song		Class	Precision	Recall	F1-Score	Support
		Podcast	9414	46		PodCast	1	.94	.97	10000
		Song	586	4954		Song	.89	.99	.94	5000

1.2 Binary Classification Models Result: Song Vs Advertisement

Table 5: Song Vs Advertisement Using Method 2(Mel-Spectrogram Applied decibel scaling to data) for English Data

Model	F1	Confusion Matrix				Classification Report				
CNN	.93		Ads	Song		Class	Precision	Recall	F1-Score	Support
		Ads	849	282		Ads	.75	.86	.80	990
		Song	141	2711		Song	.95	.91	.93	2993
VGG Net	.93		Ads	Song		Class	Precision	Recall	F1-Score	Support
		Ads	949	351		Ads	.73	.96	.83	990
		Song	41	2642		Song	.98	.88	.93	2993
Densenet	.92		Ads	Song		Class	Precision	Recall	F1-Score	Support
		Ads	836	306		Ads	.73	.84	.78	990
		Song	154	2687		Song	.95	.90	.93	2993
MobileNet	.94		Ads	Song		Class	Precision	Recall	F1-Score	Support
		Ads	678	45		Ads	.94	.68	.79	990
		Song	312	2948		Song	.90	.98	.94	2993

Table 6: Song Vs Advertisement Using Method 2 (Mel-Spectrogram Applied decibel scaling to data) for Hind Data

Model	F1	Confusion Matrix				Classification Report				
CNN	.82		Ads	Song		Class	Precision	Recall	F1-Score	Support
		Ads	635	46		Ads	.93	.64	.76	1000
		Song	365	954		Song	.72	.95	.82	1000
VGG Net	.90		Ads	Song		Class	Precision	Recall	F1-Score	Support
		Ads	796	20		Ads	.98	.80	.88	1000

(continued)

	Song	204	980

Class	Precision	Recall	F1-Score	Support
Song	.83	.98	.90	1000

Densenet — F1 .81

Confusion Matrix

	Ads	Song
Ads	563	20
Song	437	980

Classification Report

Class	Precision	Recall	F1-Score	Support
Ads	.97	.56	.71	1000
Song	.69	.98	.81	1000

MobileNet — F1 .81

Confusion Matrix

	Ads	Song
Ads	566	20
Song	434	980

Classification Report

Class	Precision	Recall	F1-Score	Support
Ads	.97	.56	.71	1000
Song	.69	.98	.81	1000

1.3 Binary Classification Models Result: PodCast Vs Advertisement

Table 7: Podcast Vs Advertisement Using Method 2 (Mel-Spectrogram Applied decibel scaling to data) for English Data

Model	F1	Confusion Matrix	Classification Report

CNN — F1 .95

Confusion Matrix

	Ads	Podcast
Ads	854	166
Podcast	136	2834

Classification Report

Class	Precision	Recall	F1-Score	Support
Ads	.84	.86	.85	990
Podcast	.95	.94	.95	3000

VGG Net — F1 .97

Confusion Matrix

	Ads	Podcast
Ads	877	76
Podcast	113	2924

Classification Report

Class	Precision	Recall	F1-Score	Support
Ads	.92	.89	.90	990
Podcast	.96	.97	.97	3000

Densenet — F1 .94

Confusion Matrix

	Ads	Podcast
Ads	712	94
Podcast	278	2906

Classification Report

Class	Precision	Recall	F1-Score	Support
Ads	.88	.72	.79	990
Podcast	.91	.97	.94	3000

MobileNet — F1 .95

Confusion Matrix

	Ads	Podcast
Ads	789	105
Podcast	201	2895

Classification Report

Class	Precision	Recall	F1-Score	Support
Ads	.88	.80	.84	990
Song	.94	.96	.95	3000

Table 8: Podcast Vs Advertisement Using Method 2(Mel-Spectrogram Applied decibel scaling to data) for Hind Data

Model	F1	Confusion Matrix	Classification Report

Model	F1	Confusion Matrix			Classification Report				
CNN	.98		Ads	Podcast	Class	Precision	Recall	F1-Score	Support
		Ads	982	121	Ads	.89	.98	.93	1000
		Podcast	18	2879	Podcast	.99	.96	.98	3000
VGG Net	.99		Ads	Podcast	Class	Precision	Recall	F1-Score	Support
		Ads	965	17	Ads	.98	.96	.97	1000
		Podcast	35	2983	Podcast	.99	.99	.99	3000
Densenet	.98		Ads	Podcast	Class	Precision	Recall	F1-Score	Support
		Ads	971	87	Ads	.92	.97	.94	1000
		Podcast	29	2913	Podcast	.99	.97	,.98	3000
MobileNet	.98		Ads	Podcast	Class	Precision	Recall	F1-Score	Support
		Ads	948	60	Ads	.94	.95	.95	1000
		Podcast	52	2940	Song	.98	.98	.98	3000

1.4 Multi Class Classification Models Result: PodCast Vs Advertisement vs Song

Table 9: Podcast Vs Advertisement vs Song Using Method 2 for English Data

Model	F1	Confusion Matrix				Classification Report				
CNN	.91		Ads	PodCast	Song	Class	Precision	Recall	F1-Score	Support
		Ads	871	298	153	Ads	.66	.88	.75	990
		PodCast	40	2647	4	PodCast	.98	.88	.93	3000
		Song	79	55	2834	Song	.95	.95	.95	2991
VGG Net	.94		Ads	PodCast	Song	Class	Precision	Recall	F1-Score	Support
		Ads	803	97	87	Ads	.81	.81	.81	990
		PodCast	87	2853	12	PodCast	.97	.95	.96	3000
		Song	100	50	2892	Song	.95	.97	.96	2991
Densenet	.87		Ads	PodCast	Song	Class	Precision	Recall	F1-Score	Support
		Ads	508	81	142	Ads	.69	.51	.59	990
		PodCast	294	2848	64	PodCast	.89	.95	.92	3000
		Song	188	71	2785	Song	.91	.93	.92	2991
MobileNet	.87		Ads	PodCast	Song	Class	Precision	Recall	F1-Score	Support
		Ads	512	128	24	Ads	.77	.52	.62	990
		PodCast	119	2674	11	PodCast	.95	.89	.92	3000
		Song	359	198	198	Song	.84	.99	.91	2991

Table 10: Podcast Vs Advertisement vs Song Using Method 2 for Hindi Data

Model	F1	Confusion Matrix				Classification Report				
CNN	.92		**Ads**	**PodCast**	**Song**	Class	Precision	Recall	F1-Score	Support
		Ads	578	14	78	Ads	.86	.58	.69	1000
		PodCast	30	2942	23	PodCast	.98	.98	.98	3000
		Song	392	44	2899	Song	.87	.97	92	3000
VGG Net	.94		**Ads**	**PodCast**	**Song**	Class	Precision	Recall	F1-Score	Support
		Ads	660	14	24	Ads	.94	.66	.78	1000
		PodCast	15	2965	23	PodCast	.99	.99	.99	3000
		Song	325	19	2953	Song	.90	.98	.94	3000
Densenet	.89		**Ads**	**PodCast**	**Song**	Class	Precision	Recall	F1-Score	Support
		Ads	711	55	275	Ads	.68	.71	.70	1000
		PodCast	47	2932	114	PodCast	.95	.98	.96	3000
		Song	242	13	2611	Song	.91	.87	.89	3000
MobileNet	.88		**Ads**	**PodCast**	**Song**	Class	Precision	Recall	F1-Score	Support
		Ads	539	120	101	Ads	.71	.54	.61	1000
		PodCast	23	2822	38	PodCast	.98	.94	.96	3000
		Song	438	58	2861	Song	,85	.95	.90	3000