Final Project Submission

Aditya DSS & Aritra Deb

Which ML models did you train for the task ?

Ans. On our sound classification task we have used ANN,RNN,CNN and YamNet.

1 Compare the models on the basis of

a) Complexity, Ease of interpretation

- **Artificial Neural Network (ANN):** The ANN model is relatively easier to interpret, providing a balance between simplicity and effectiveness. Notably, it boasts a modest memory size of only 50 KB.
- **Convolutional Neural Network (CNN):** The CNN model, while achieving high performance, comes with increased complexity and may pose challenges in interpretability. Additionally, its memory size is substantial, nearing a quarter of a gigabyte.
- **Recurrent Neural Network (RNN):** Similar to the CNN, the RNN model exhibits complexity but performs well in classifying audio signals. However, specific memory size details are not provided in the given information.

(b) Accuracy or other relevant performance metrics
Ans. Here are the tables of model performance.

ANN Model				
Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
1	4.6488	0.5268	4.3846	0.5000
2	2.7308	0.6518	0.8677	0.6250
3	2.4003	0.5268	1.3934	0.6250
4	2.1701	0.6607	1.8024	0.5000
5	1.1722	0.6696	0.5749	0.8750
6	1.4174	0.6696	2.1183	0.5000
7	0.9078	0.7321	0.6455	0.8750
8	0.8263	0.7232	1.1917	0.5000
9	0.6638	0.7500	0.4081	0.8750
10	0.5173	0.7946	0.5946	0.7500
11	0.3891	0.8304	0.2425	0.8750
12	0.3202	0.8482	0.5484	0.6250
13	0.3939	0.8304	0.1887	0.8750
14	0.3797	0.8214	0.9087	0.6250
15	0.4627	0.8036	0.2520	0.8750
16	0.4224	0.8214	0.3362	0.8750
17	0.3613	0.8214	0.4744	0.7500
18	0.3053	0.8661	0.2187	0.8750
19	0.3125	0.8571	0.4257	0.8750
20	0.3046	0.8839	0.2350	0.8750
21	0.2899	0.8393	0.5813	0.6250
22	0.3445	0.8304	0.3978	0.8750
23	0.2772	0.8929	0.2505	0.7500
24	0.3081	0.8839	0.4587	0.8750
25	0.2718	0.9107	0.2304	0.8750
26	0.2769	0.8661	0.3400	0.8750
27	0.2342	0.9018	0.2441	0.8750
28	0.1869	0.9286	0.4856	0.7500
29	0.2244	0.8750	0.3391	0.8750
30	0.1989	0.9375	0.2283	0.8750
31	0.2418	0.8929	0.5964	0.7500
32	0.2050	0.9107	0.2301	0.8750
33	0.3174	0.8571	0.5278	0.8750
34	0.3626	0.8125	0.2160	0.8750
35	0.2873	0.8661	0.1823	0.8750
36	0.2340	0.9286	0.4252	0.8750
37	0.1930	0.9286	0.2826	0.8750
38	0.2277	0.9018	0.4394	0.7500
39	0.1811	0.9286	0.1859	0.8750
40	0.1643	0.9375	0.5425	0.7500
Total training time: 3.71 seconds				

Table 1: ANN Model Training and Validation Metrics Over Epochs

RNN Model				
Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
1	0.6892	0.4911	0.6721	0.3750
2	0.6723	0.5714	0.7134	0.5000
3	0.6555	0.5893	0.6944	0.5000
4	0.6428	0.6786	0.6298	0.6250
5	0.6231	0.7054	0.5665	1.0000
6	0.5922	0.7321	0.4628	1.0000
7	0.5559	0.7411	0.3443	1.0000
8	0.5054	0.7768	0.2986	1.0000
9	0.4770	0.7768	0.2950	0.8750
10	0.4632	0.7500	0.2648	0.8750
11	0.4278	0.7857	0.1872	1.0000
12	0.4207	0.7946	0.1710	1.0000
13	0.3985	0.8036	0.1265	1.0000
14	0.3726	0.8304	0.1198	1.0000
15	0.3484	0.8304	0.0865	1.0000
16	0.3219	0.8661	0.0792	1.0000
17	0.3004	0.8661	0.0865	1.0000
18	0.2988	0.8571	0.0971	1.0000
19	0.3122	0.8750	0.0517	1.0000
20	0.3493	0.8393	0.0392	1.0000
21	0.3058	0.8571	0.0703	1.0000
22	0.2527	0.8929	0.0642	1.0000
23	0.3008	0.8929	0.0689	1.0000
24	0.2678	0.8839	0.0527	1.0000
25	0.2724	0.8661	0.0593	1.0000
26	0.2452	0.8839	0.0408	1.0000
27	0.2419	0.9018	0.0345	1.0000
28	0.2255	0.9018	0.0317	1.0000
29	0.2222	0.8839	0.0291	1.0000
30	0.2032	0.9018	0.0211	1.0000
31	0.1961	0.9018	0.0251	1.0000
32	0.1945	0.9018	0.0179	1.0000
33	0.2190	0.8929	0.0286	1.0000
34	0.2243	0.8839	0.0295	1.0000
35	0.1784	0.9286	0.1183	0.8750
36	0.2348	0.8839	0.1582	0.8750
37	0.2385	0.8750	0.1736	0.8750
38	0.2450	0.8750	0.0770	1.0000
39	0.1956	0.9196	0.0352	1.0000
40	0.2026	0.9018	0.0287	1.0000
Total training time: 15.40 seconds Test Accuracy: 93.75%				

Table 2: RNN Model Training and Validation Metrics Over Epochs

YAMNET Model				
Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
1	0.7997	0.5625	1.7028	0.5000
2	0.5379	0.7344	2.3042	0.8125
3	0.4269	0.8594	2.6696	0.8750
4	0.3982	0.8438	2.9609	0.8125
5	0.3036	0.8750	3.2713	0.8750
6	0.2653	0.9219	3.5789	0.8750
7	0.2259	0.9531	3.8964	0.8125
8	0.2093	0.9531	4.2072	0.8125
9	0.2135	0.9531	4.5146	0.8125
10	0.1691	0.9688	4.7585	0.8125
1/1 [==========] - 0s 32ms/step - loss: 4.7585 - accuracy: Validation Loss: 4.7585, Validation Accuracy: 81.25%				

Table 3: YAMNET Model Training and Validation Metrics Over Epochs

CNN Model					
Epoch	Training Loss	Training Recall	Training Precision	Validation Loss	Validation Recall
1	15.8019	0.7059	0.6857	1.6755	0.9333
2	1.6292	0.9615	0.7143	0.5449	0.9333
3	0.5452	0.9643	0.8182	0.3454	0.8000
4	0.4414	0.9231	0.9231	0.2029	0.9231

Table 4: Training and Validation Metrics Over Epochs. In CNN we getting Accuracy: 1.0 Precision: 1.0, Recall: 1.0, F1 Score: 1.0. Our Training Time is **17 Seconds**

c) **Variance of the implemented models:**
Artificial Neural Network (ANN): The fluctuation in the validation accuracy and the increasing validation loss suggest a certain level of variance. The model could be sensitive to changes in the training data, potentially indicating overfitting.
YAMNet: Similar to the ANN, there are signs of potential overfitting as seen in the increasing validation loss. This may imply some variance in the model's ability to generalize to new data.
Recurrent Neural Network (RNN): The RNN also exhibits signs of potential overfitting, with the validation loss increasing slightly. This suggests that the model might be sensitive to variations in the training data.)

Conclusion

After training and evaluating multiple machine learning models for sound classification, we can draw the following conclusions:

- **Artificial Neural Network (ANN):** Utilizing a structure with three hidden layers of widths 64, 128, and 32 neurons, the ANN model exhibits competitive performance. However, further analysis of its complexity and interpretability is required.
- **Recurrent Neural Network (RNN):** The RNN model demonstrates promising results with a test accuracy of 93.75%. Its training time is comparatively longer, taking approximately 15.40 seconds.
- **YAMNET Model:** While the YAMNET model achieves an 81.25% validation accuracy, it is noteworthy that the model's performance is accompanied by a total training time of 3.71 seconds.
- **Convolutional Neural Network (CNN) :** The CNN model stands out as the best-performing model, showcasing high accuracy and precision. However, it comes with a substantial computational cost due to its large number of parameters—111,352,605 (424.78 MB) in total, making it a resource-intensive option.

Compare your results with those mentioned in the research papers. Identify where there is scope for improvement.

In our evaluation, we have faithfully implemented the data augmentations proposed in our research paper, achieving a 100% accuracy with one of our models, specifically the Convolutional Neural Network (CNN). However, it is crucial to acknowledge the limitations of our study. The selected dataset, while representative, exhibits linearity and a relatively small number of data points. Despite these challenges, our models demonstrate commendable performance, showcasing the potential for even greater success with a more diverse and expansive dataset. Acquiring additional data emerges as a paramount task, as it can significantly enhance the robustness and generalization capabilities of our models. Furthermore, addressing the computational cost associated with our CNN model is imperative. Exploring regularization techniques and incorporating strategies like label smoothing can contribute to refining our models, fostering a more nuanced understanding of audio signal classification while ensuring scalability and efficiency in real-world applications.