Data Classification with Mudestreda Multimodal Device State Recognition Dataset & Real Industrial Milling Device data in Time Series and Spectral Images

Device State Classification with Images and Dynamometer Data

Group 24: **Eric** Bressinger, **Nikhil** Gandudi Suresh, **Sharvari** Deshmukh

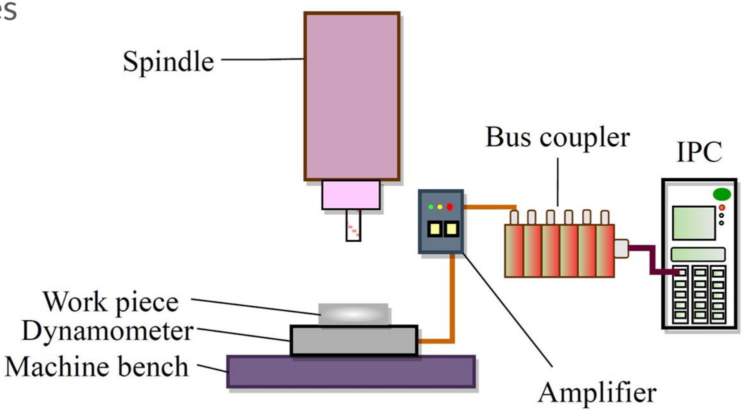
Introduction

Motivation: The combination of data from different modalities offers complementary information leading to better understanding.

Objective: Use image and force data to predict if the blade/tool is:

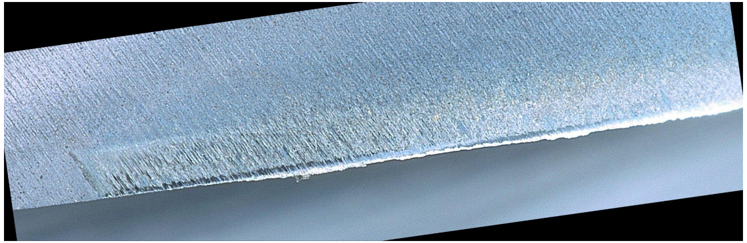
a. Sharp
b. Dull
c. Used (and worn off)

with as small a model as possible.

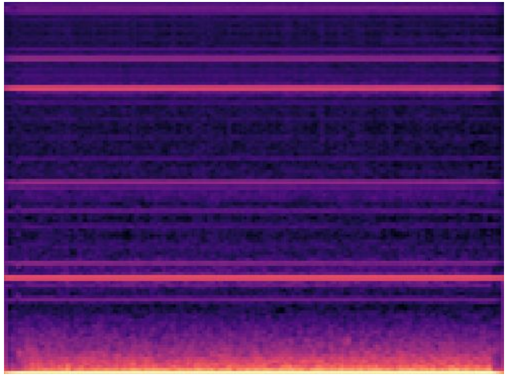


Truchan/Mudestreda

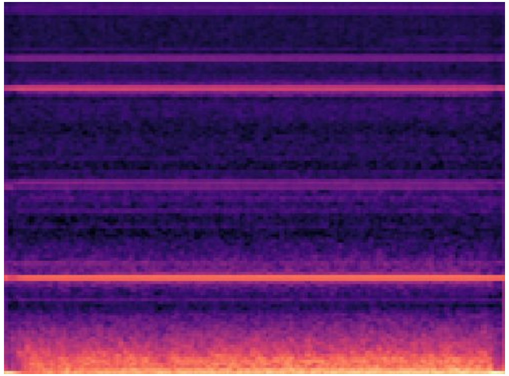
Datasets



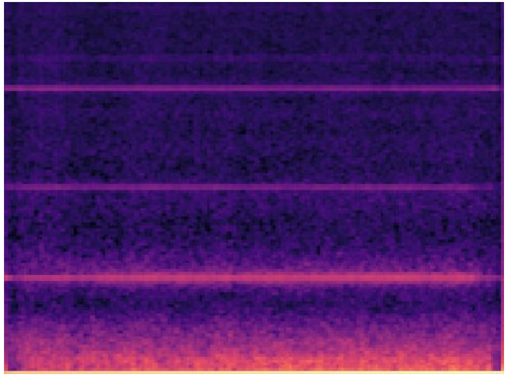
Tool Image



SpecX



SpecY



SpecZ

Literature review

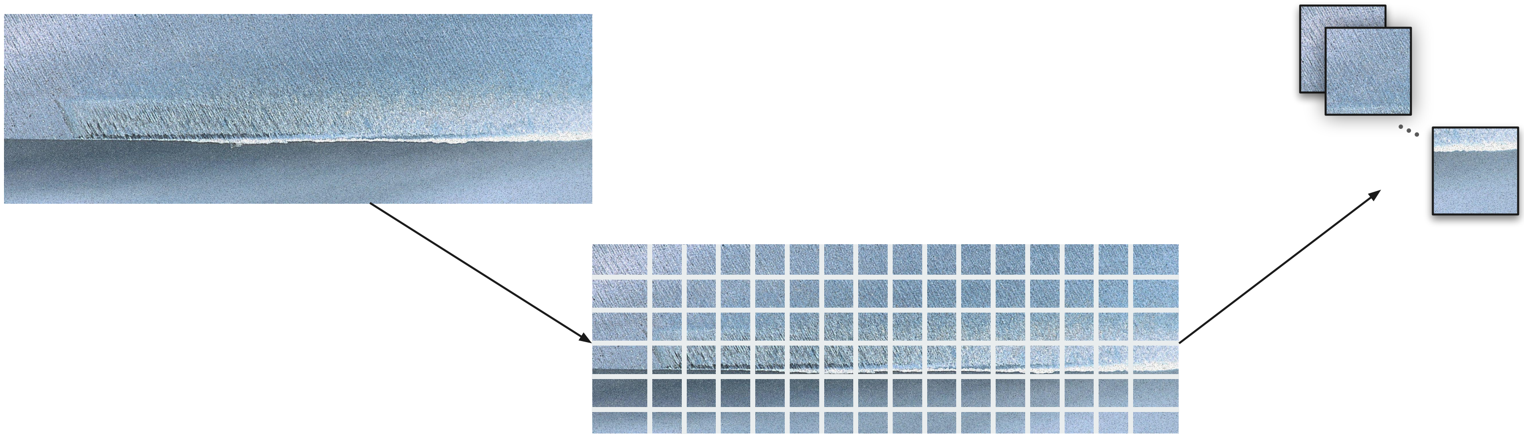
- Tadas Baltrusaitis, Chaitanya Ahuja, and Louis-Philippe Morency. 2019. Multimodal Machine Learning: A Survey and Taxonomy. IEEE Trans. Pattern Anal. Mach. Intell. 41, 2 (February 2019), 423–443. https://doi.org/10.1109/TPAMI.2018.2798607

- Truchan, Hubert, et al. "Multimodal Isotropic Neural Architecture with Patch Embedding." International Conference on Neural Information Processing. Singapore: Springer Nature Singapore, 2023.

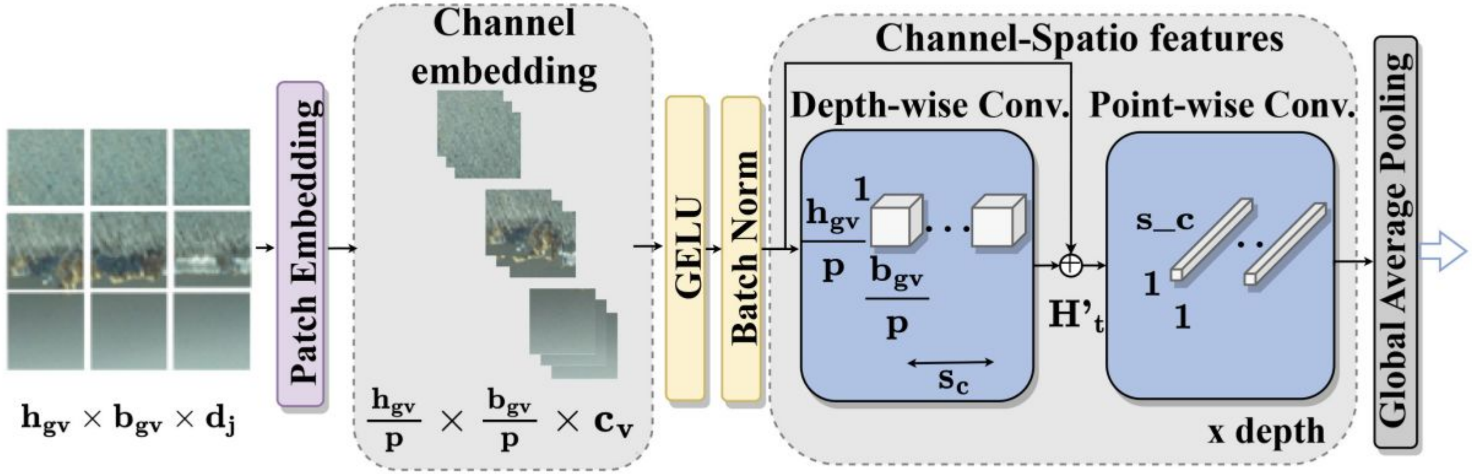
Methodology

Patch embeddings enable in capturing **global** dependencies. Thus, helping in image understanding and analysis. However, we **don't** use transformer.

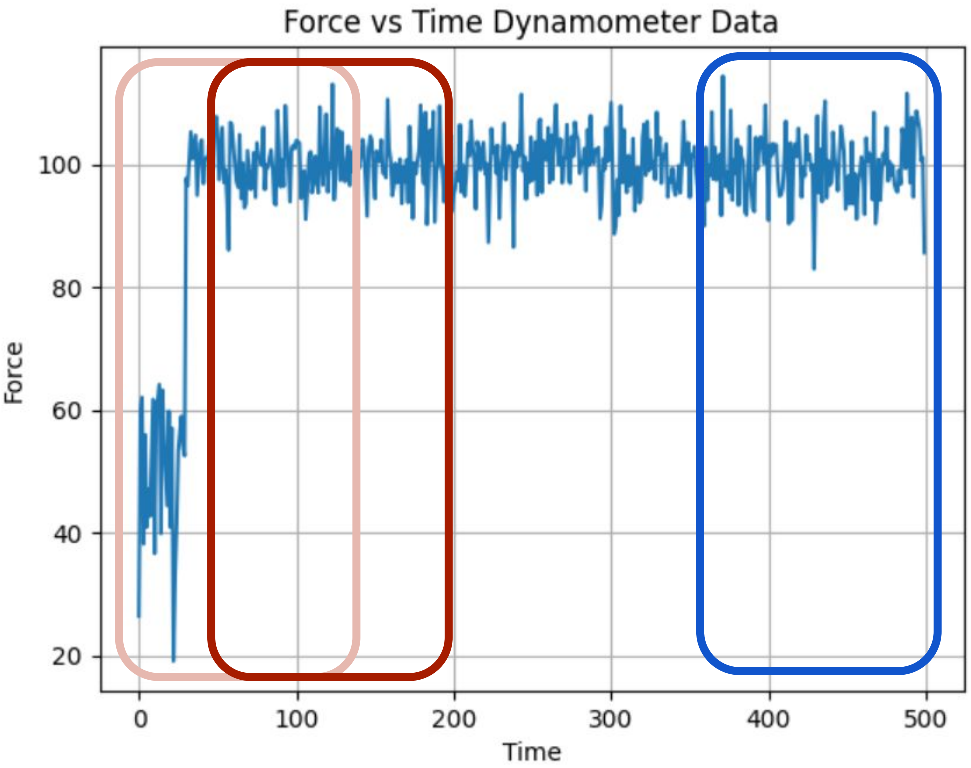
We use independent models for each modality and concatenate the results from each modality and feed it into a mixer model, which analyses and gives a final classification result.

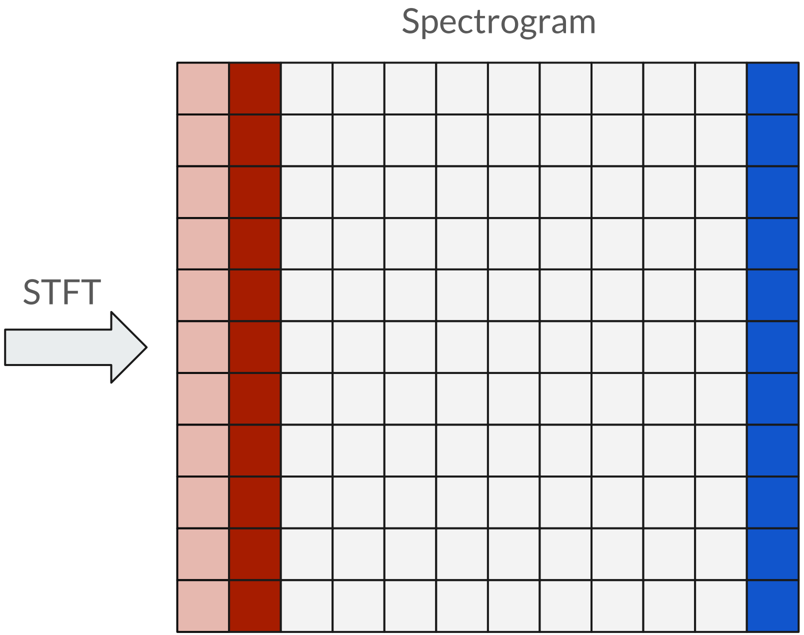


Visual Modality Pathway

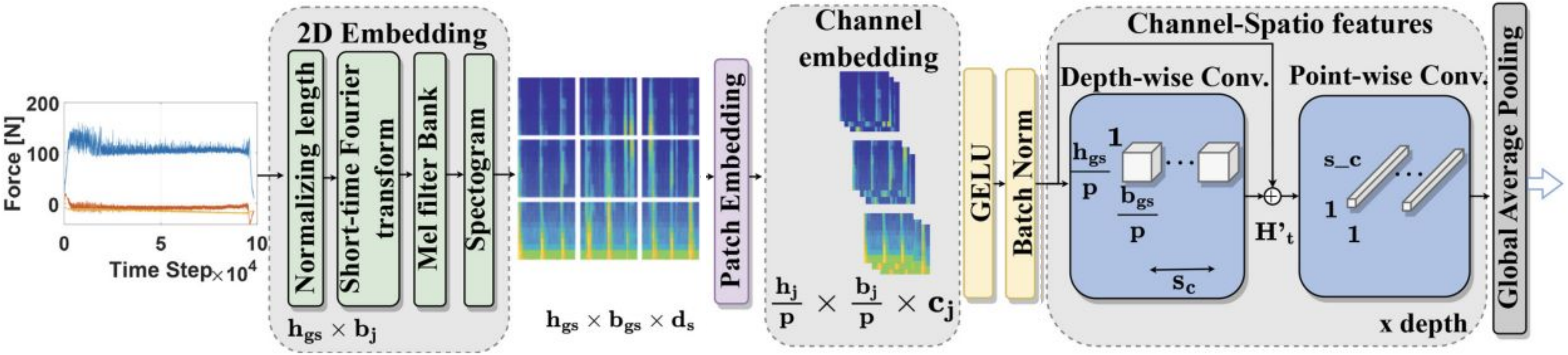


Time Series Modality Pathway

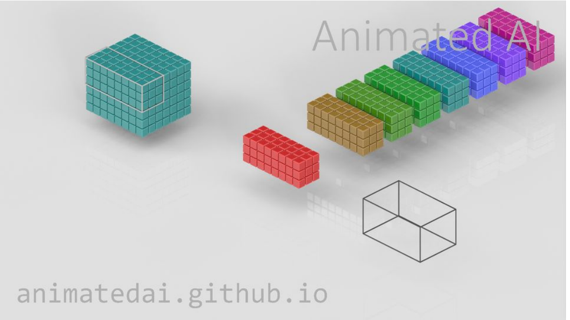


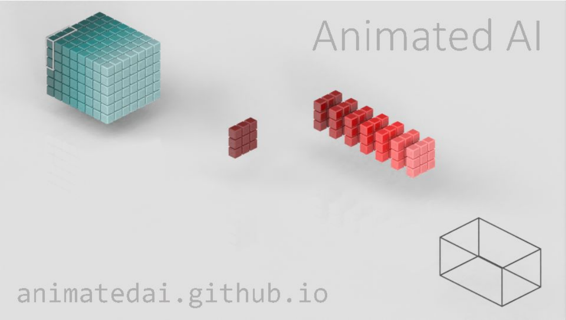


Time Series Modality Pathway



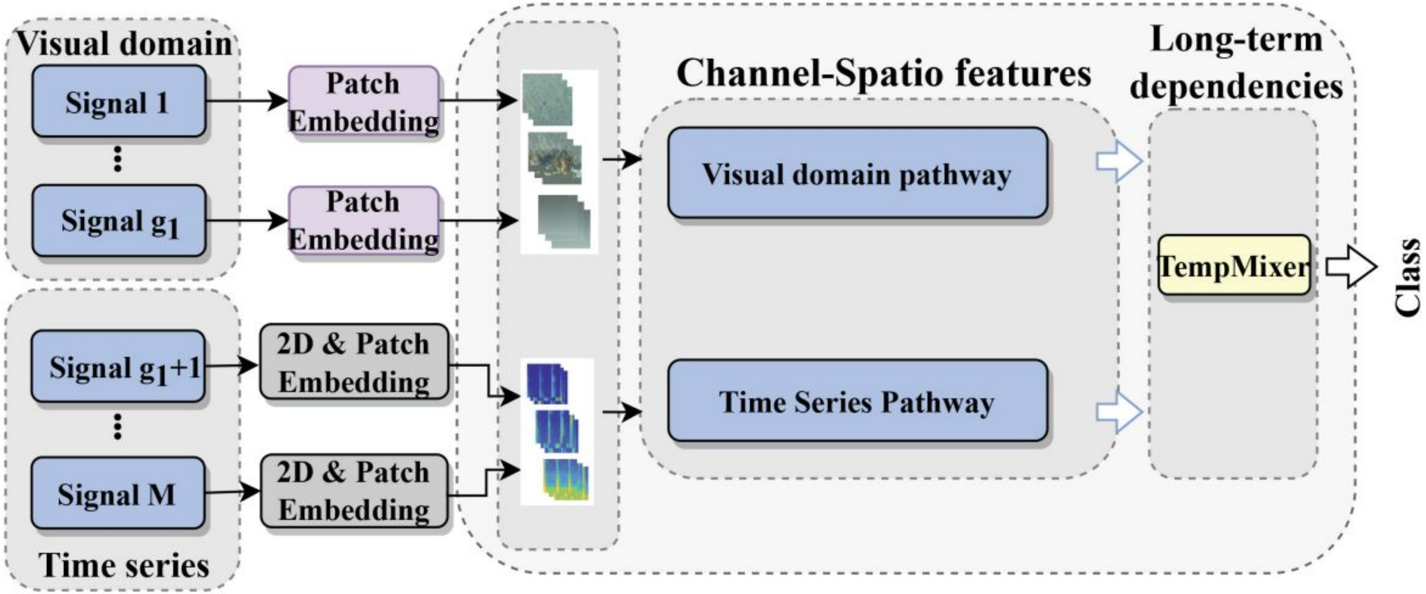
Conv2D





Depth-wise Convolution

Multimodal architecture



Data distribution

- **3072** Samples
- **4** Modalities
- **3** Classes

- Original data vs Augmented data
- Random distribution vs Tool Distribution

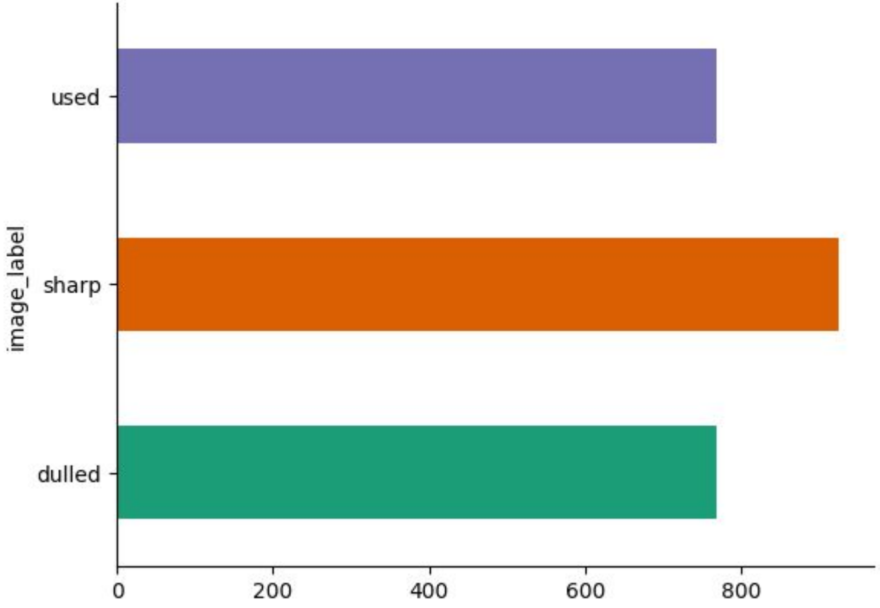
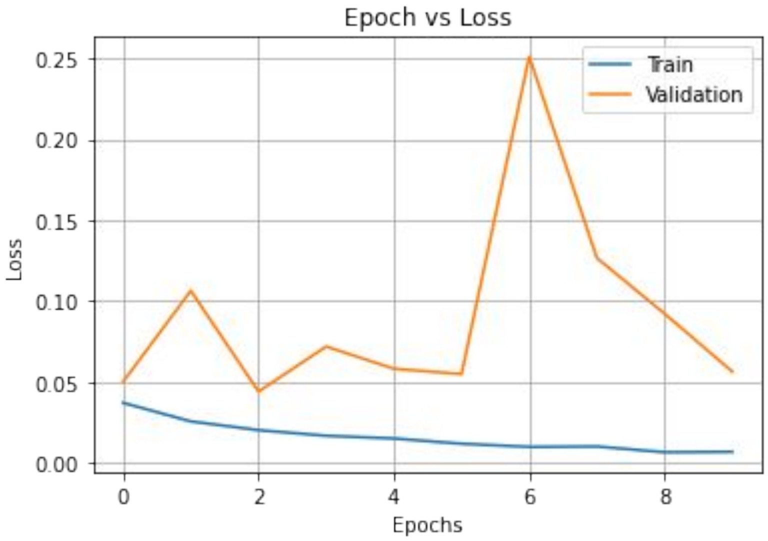


Image Data Results [Original Data]

- **Test Accuracy: 43.75%**
- **Model size: 9.53 MB**

(Adam optimizer: learning rate = 0.001, weight decay = 0.0001, batch size = 32, epochs = 10)



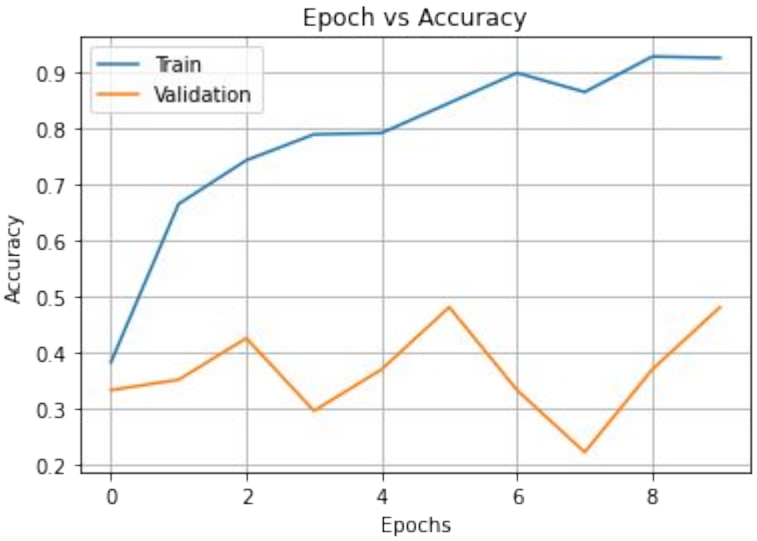
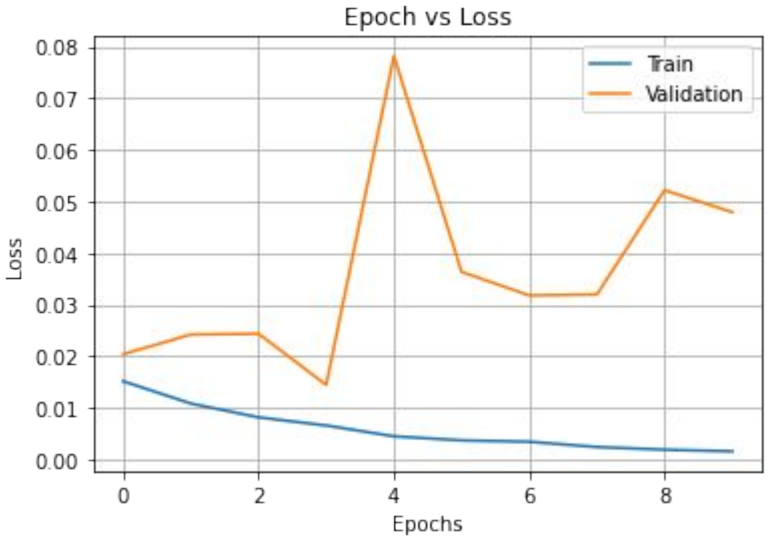
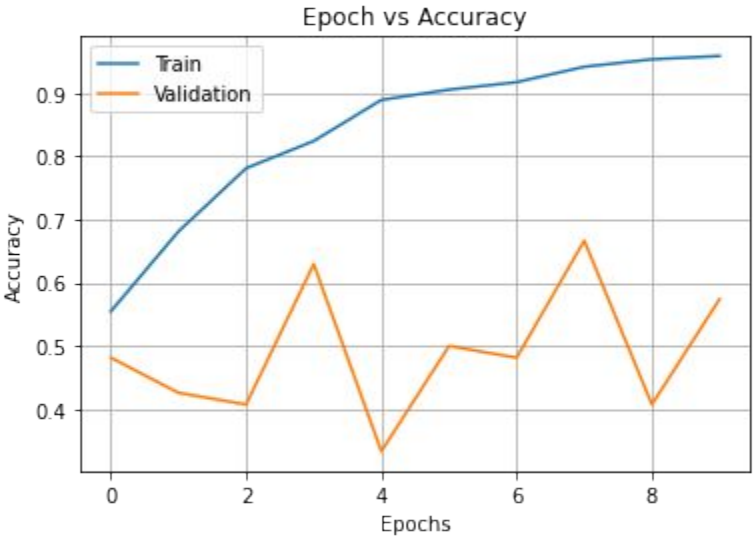


Image Data Results [Augmented Data]

- **Test Accuracy: 62.5%**
- **Model size: 9.53 MB**

(Adam optimizer: learning rate = 0.001, weight decay = 0.0001, batch size = 32, epochs = 10)

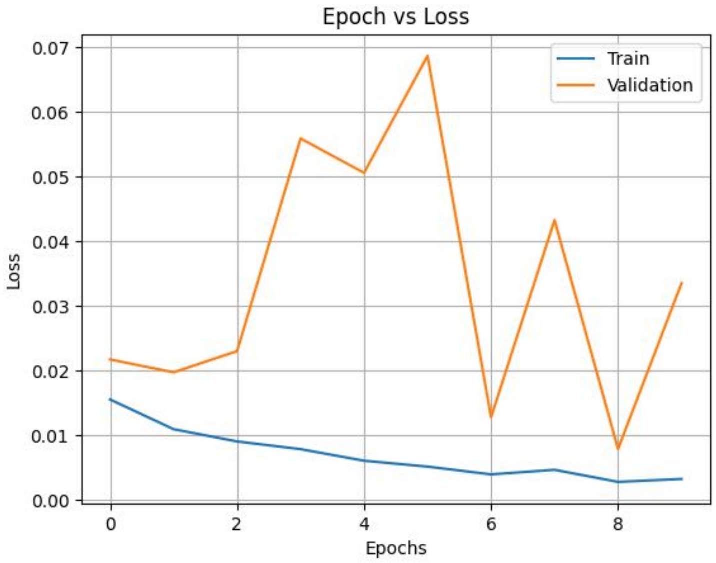


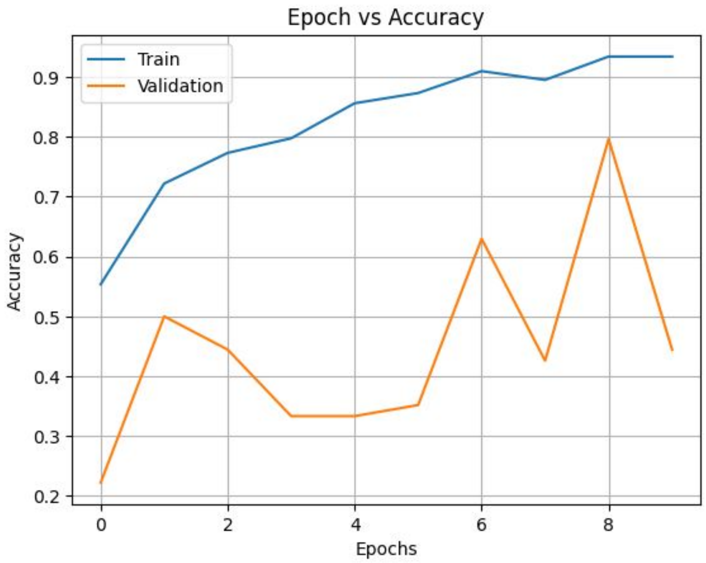


Time Series Data Results [Original Data]

- **Test Accuracy: 72.91%**
- **Model size: 5.05 MB**

(Adam optimizer: learning rate = 0.001, weight decay = 0.0001, batch size = 64, epochs = 10)

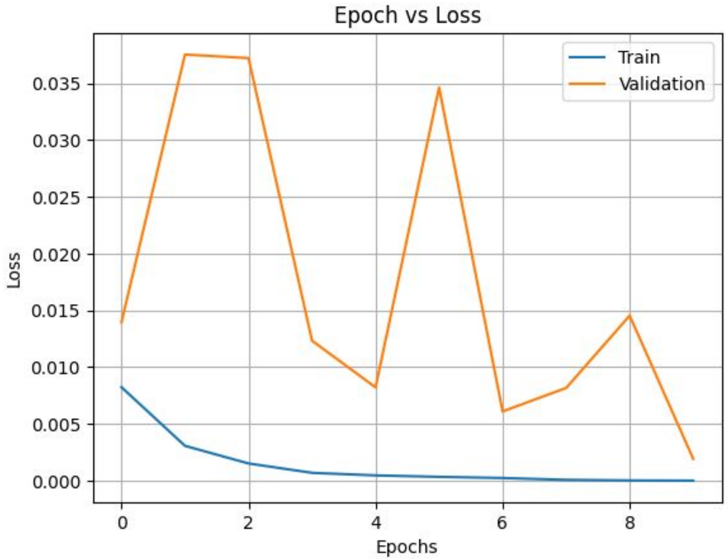


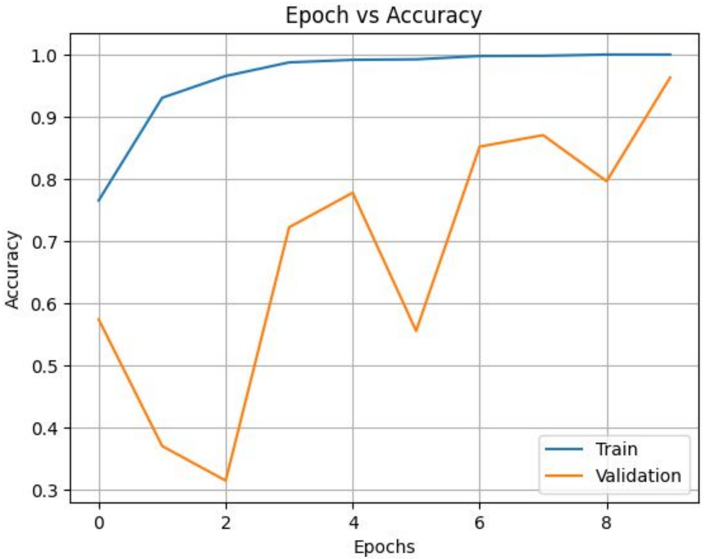


Time Series Data Results [Augmented Data]

- **Test Accuracy: 97.9%**
- **Model size: 5.05 MB**

(Adam optimizer: learning rate = 0.001, weight decay = 0.0001, batch size = 64, epochs = 10)

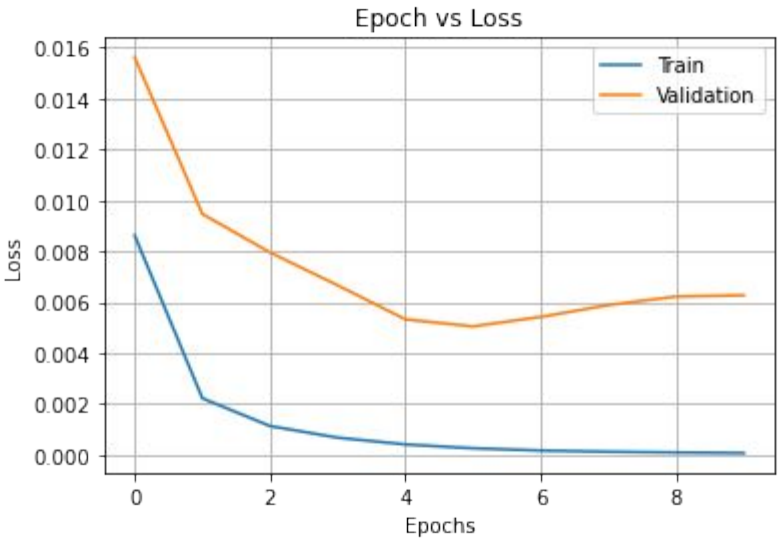


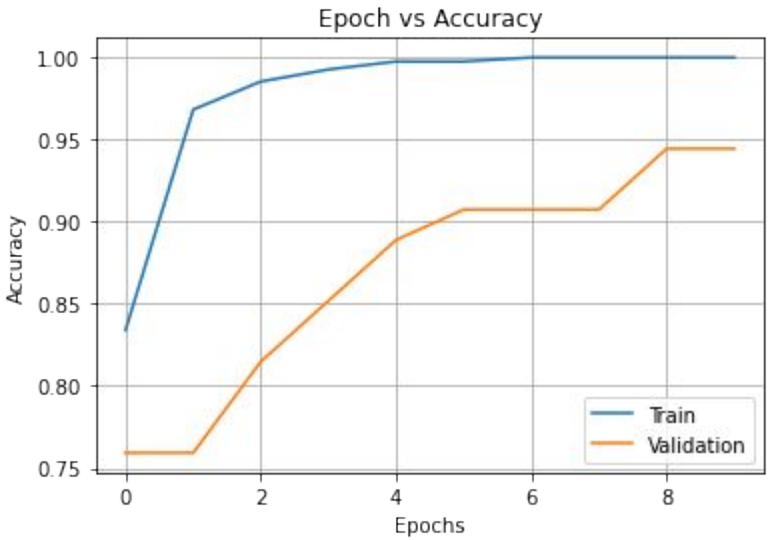


Multimodal Results [Original Data]

- **Test Accuracy: 87.5%**
- **Model size: 17.3 MB**

(Adam optimizer: learning rate = 0.001, weight decay = 0.0001, batch size = 64, epochs = 10)

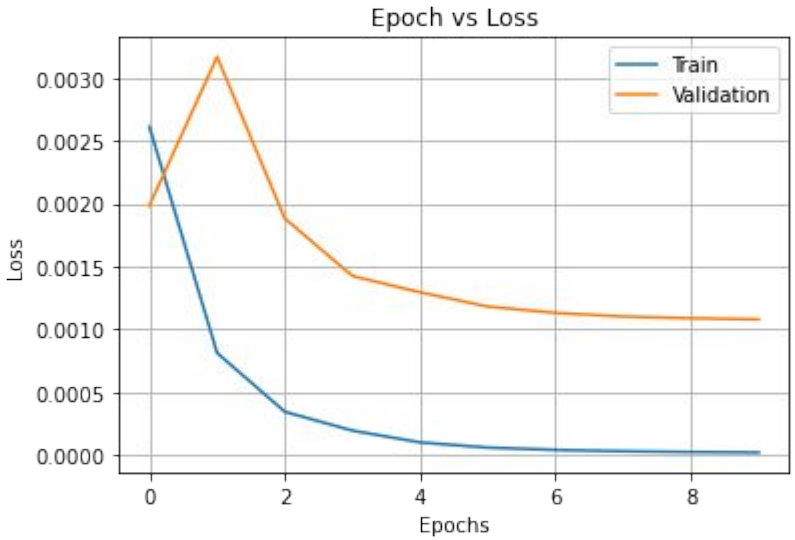


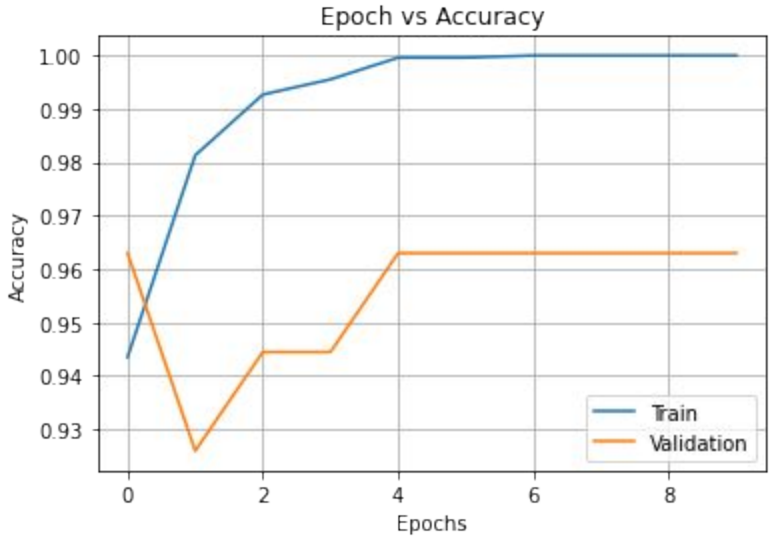


Multimodal Results [Augmented Data]

- **Test Accuracy: 95.83%**
- **Model size: 17.3 MB**

(Adam optimizer: learning rate = 0.001, weight decay = 0.0001, batch size = 64, epochs = 10)





Conclusion and discussion

- **Efficient** way to combine multi-modal data for prediction or regression.
- Depth-wise convolutions **reduce computation and size** of the model.
- Model with Conv2d → **~200MB**, whereas model with Depthwise Conv2d (followed by pointwise convolution) → **~15MB**
- Model size grows **linearly** with increasing modalities.
- Efficient way to run models on **edge devices**.
- Easily **parallelizable** with each modality running on its own chain.

Future work

- Test our model on open source **benchmark dataset - MeX** (a human activity recognition dataset containing time series data (acceleration, proximity, pressure) and images) **[In-Progress]**
- Include **additional modalities** like acoustic emission by the machinery, power consumption, etc.
- Extend the work for **Remaining Useful Life (RUL) estimation** posed as a regression problem, considering the final observation of each tool as the endpoint of its lifecycle.

References

Chen L, Li S, Bai Q, Yang J, Jiang S, Miao Y. Review of Image Classification Algorithms Based on Convolutional Neural Networks. Remote Sensing. 2021; 13(22):4712. https://doi.org/10.3390/rs13224712

"Mudestreda Multimodal Device State Recognition Dataset." Papers with Code, 2023, https://paperswithcode.com/dataset/mudestreda.

Shorten, C., Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. J Big Data 6, 60 (2019). https://doi.org/10.1186/s40537-019-0197-0.

Tadas Baltrusaitis, Chaitanya Ahuja, and Louis-Philippe Morency. 2019. Multimodal Machine Learning: A Survey and Taxonomy. IEEE Trans. Pattern Anal. Mach. Intell. 41, 2 (February 2019), 423–443. https://doi.org/10.1109/TPAMI.2018.2798607

Truchan, Hubert, and Zahra Admadi. *Mudestreda Multimodal Device State Recognition Dataset*. Zenodo, 24 Jan. 2024, https://doi.org/10.5281/zenodo.8238653.