PERFORMANCE OF MULTILAYER PERCEPTRONS ON ECG ANALYSIS

CS615: DEEP LEARNING

DAVID EVERLY

6/8/2025

BACKGROUND

- US Healthcare System substantially behind in basic safety[1]
- Annually, 98,000 deaths due to medical errors in US hospitals[1]
- IOM encouraged use of IT[2]
  - Simplify workflow processes
  - Reduce waste
  - Improve healthcare access

1 (INSTITUTE OF MEDICINE [IOM], 2000, P.5)
2 (IOM, 2009)

BACKGROUND

- Effective treatment of Acute Coronary Syndrome (ACS) is time sensitive
- Current guidelines identify ACS on ECG by:
  - ST-segment elevation (STEMI)
  - Elevated biomarker detection (NSTEMI)
- 24-35% of those with NSTEMI need emergent catheterization
  - Delayed while awaiting biomarker elevation due to inability to detect on ECG

1 (AL-ZAITI ET AL., 2023)

PREVIOUS WORK

- Literature Review 2022
  - 59 Deep Learning studies analyzing ECG data
  - Convolutional Neural Networks and ResNet best performing
    - Reported over 97% accuracy
  - ECG data is noisy
    - May be more suited to DL than ML

1 (XIONG ET AL., 2022)

PREVIOUS WORK

- 2023 prospective, observational
  - 2023 prospective, observational cohort
  - 12-Lead ECG Data
  - 10 different ML classifiers
  - Random Forest Model achieved AUROC 0.91(95% CI = 0.87-0.96)
    - Outperformed practicing clinicians and commercial ECG systems

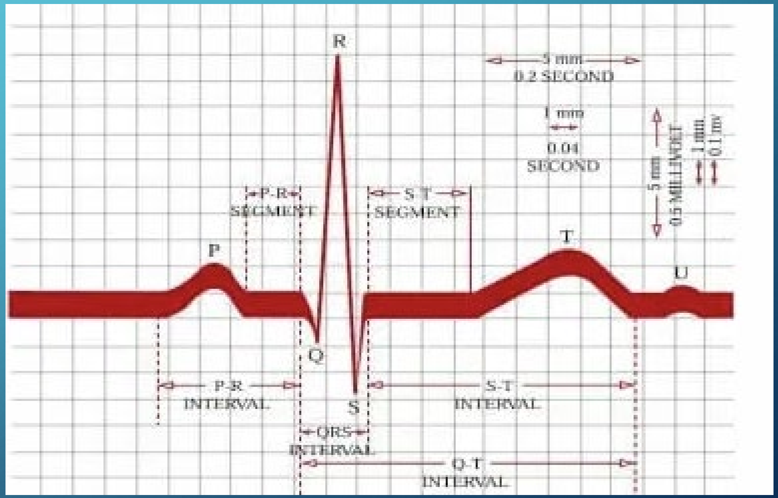
1 (AL-ZAITI ET AL., 2023)

PURPOSE

- To analyze the performance of MLP classification models on identifying abnormal heart rhythms and myocardial ischemia on ECG data at various degrees of complexity

QUESTION

- Can a MLP classification model accurately identify cardiac dysrhythmias and myocardial ischemia from ECG tracings?



DATA

- ECG 200[1]
  - N = 200
  - 1D time series of ECG tracing, single cycle
  - Classes: Normal vs Abnormal
- MIT-BIH[3]
  - N = 47 subjects
  - 2 lead ECG tracing, 30 minutes long
  - Classes: Normal vs Ischemia
- ECG 5000[2]
  - N = 5000
  - 1D time series of ECG tracing, single cycle
  - Classes:
    - Normal
    - R-on-T PVC
    - PVC
    - Escape Beat
    - Ventricular Fusion

1 (DAU, ET AL., 2018)
2 (GOLDBERGER ET AL., 2000)

3 (MOODY & MARK, 2001)

DATA

- Compare our model against Al-Zaiti's 2023 RF Model
  - 12-Lead ECG Data
  - Random Forest Model achieved AUROC 0.91(95% CI = 0.87-0.96)
    - AUROC 0.91 = Benchmark for our Model

1 (AL-ZAITI ET AL., 2023)

PRE-PROCESSING

- ECG 200 and ECG5000 data
  - Available from author cleaned and split 50/50 training/validation sets
  - Data were re-combined, shuffled, and split 75/25 training/validation
  - ECG5000 class outcomes were one-hot-encoded
    - Significant oversampling of normal tracings

PRE-PROCESSING

- MIT-BIH data
  - Much more complicated; 2 lead ECG tracing over time
  - Sliding window was created to take 6 second snapshot of tracing
  - Any observation with missing data was omitted
  - Each lead data within a window was concatenated together to form single observation
    - 3000 features per observation
  - Subsample of 1725 observations; 861 normal and 864 ischemia

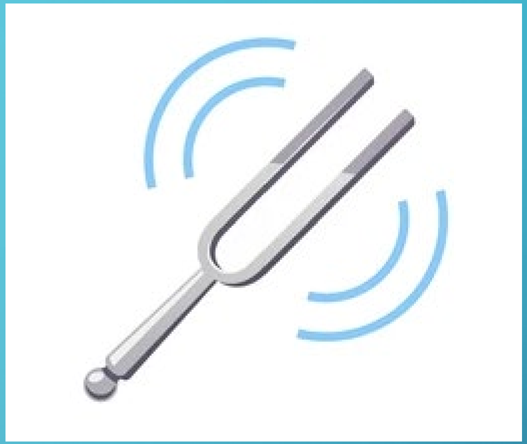
PRE-PROCESSING

- Full ECG data (Al-Zaiti)
  - Data features taken AS-IS
    - No randomization or shuffling. We used their training/validation data, as published
    - N = 7313 -> Training = 4026 , Validation = 3287
      - 74 features, including patient demographics
    - BUT very imbalanced distribution, very high majority of class normal
      - Approximately 370 of each class in both training and validation set
    - Missingness is a problem
      - Extreme outliers (> 5 SD) were removed
      - All NaN values were replaced by column mean after removing extreme outliers

METHODS

- On each dataset, compare performance of
  - Shallow MLP (5 Layers); INPUT -> FCC -> TANH -> LOGSIG -> LOG LOSS
  - Deep MLP (22 Layers); INPUT -> (FCC -> TANH) x 9 -> FCC -> LOGSIG -> LOGLOSS
  - Deep MLP with Skip residuals
    - Same 22 Layers as Deep but each FCC->TANH sequence is wrapped with skip residuals and batch normalization layer
    - FCC use hidden state size funneling ; dimension size -2 every other skip layer block
    - TANH is prone to dead activations, skip layer should attenuate that problem
  - TANH proved too unstable for Full ECG data, so we switched all TANH to ReLU
    - BATCHNORM layer just before final LOGSIG for added stability

METHODS



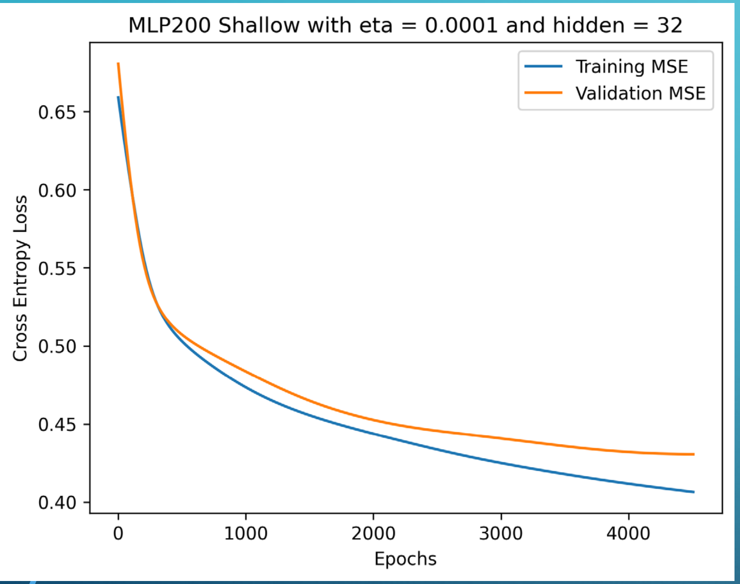
- Hyperparameter Tuning
  - Hidden State (FCC output dimension) and learning rate
    - Optimized by painful trial and error for each model
      - Iterative breadth-search
      - All permutations of Eta = [10^{-3},10^{-4},10^{-5}] and Hidden State Size [8,16,32,64]
    - ECG200 -------------------------> Eta = .00001, Hidden State Size = 32
    - ECG5000 ------------------------> Eta = .0001, Hidden State Size = 32
    - ECG5000 Multiclass -----------> Eta = .0001, Hidden State Size = 32
    - ECG5000 Subsampled -------> Eta = .00001, Hidden State Size = 32
    - MIT-BIH --------------------------> Eta = .00001, Hidden State Size = 32
    - FULL ECG ------------------------> Eta = .00032, Hidden State Size = 16
  - All FCC use ADAM except the FULL ECG pipeline
    - Removed due to instability

METHODS

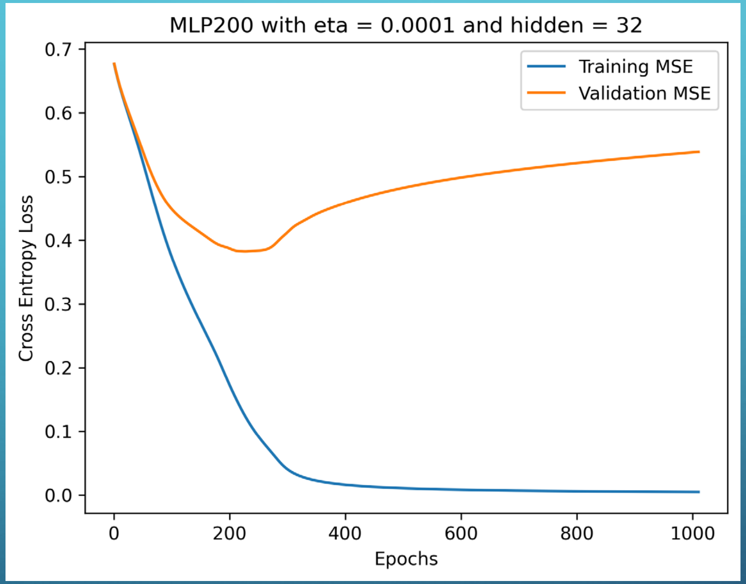
- FCC weight and bias initialization
  - Random, uniformly distributed, Xavier
  - $\left[-\sqrt{\frac{6}{h_{in}+h_{out}}},\sqrt{\frac{6}{h_{in}+h_{out}}}\right]$

RESULTS-ECG200

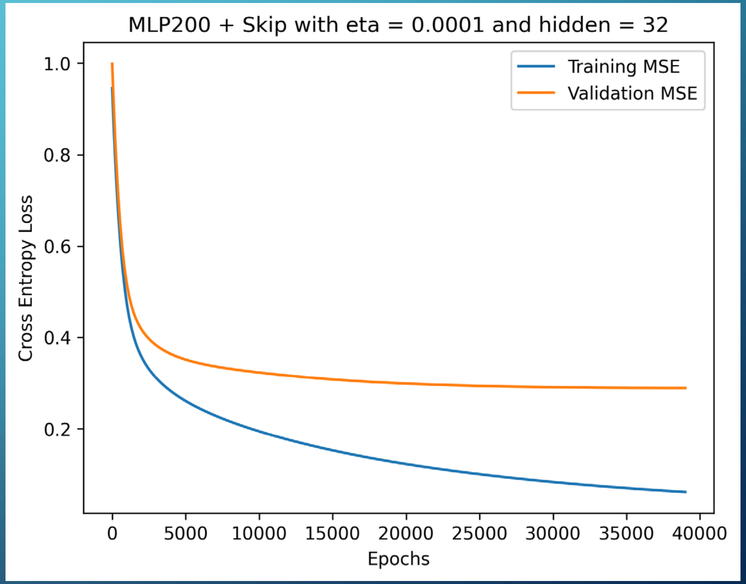
Shallow - Trained Well



Deep – Exploding Gradient



Skip Residuals Fix it!
Still overfit

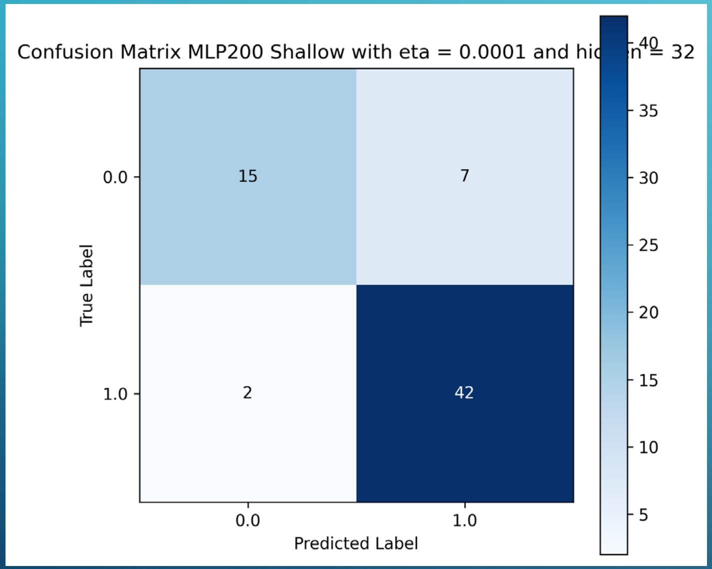


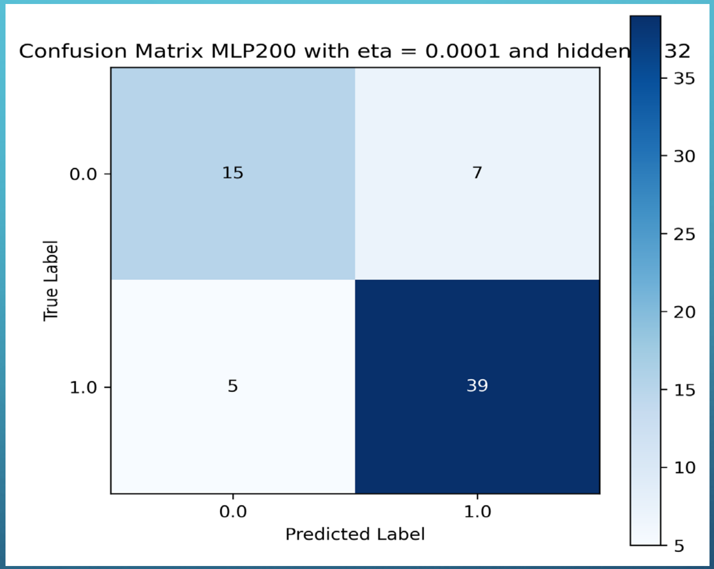
RESULTS-ECG200

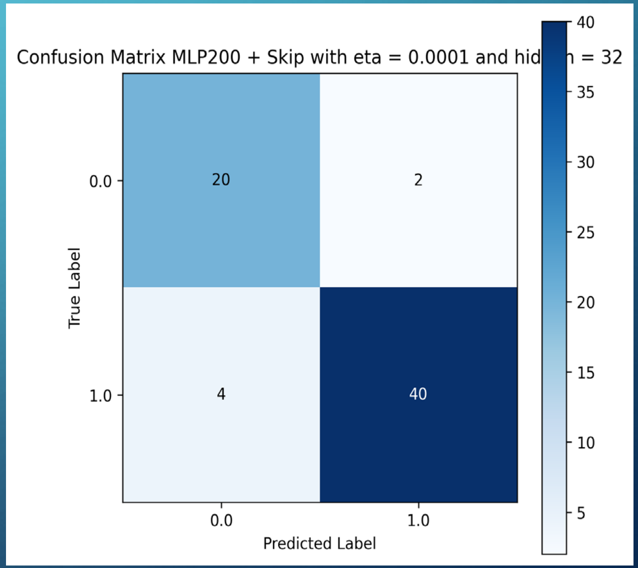
Shallow Accuracy = 86%

Deep Accuracy = 81%

Deep with Skip Accuracy = 91%

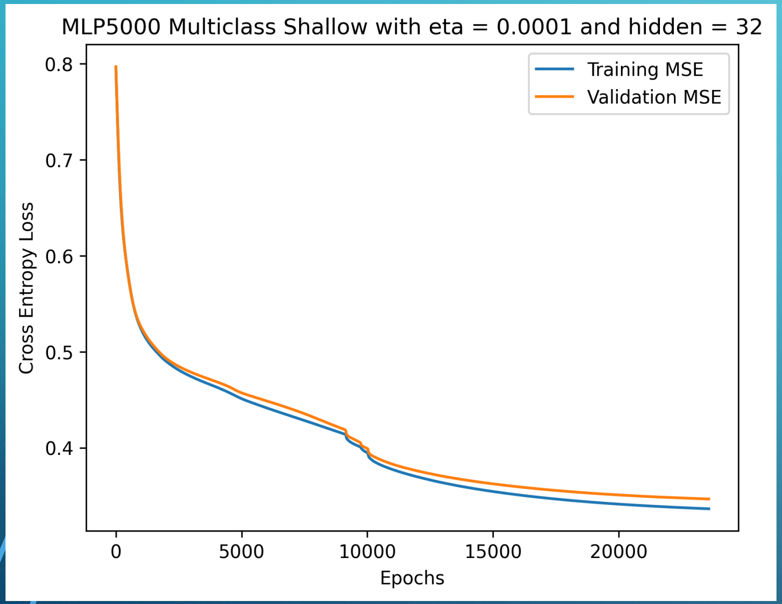




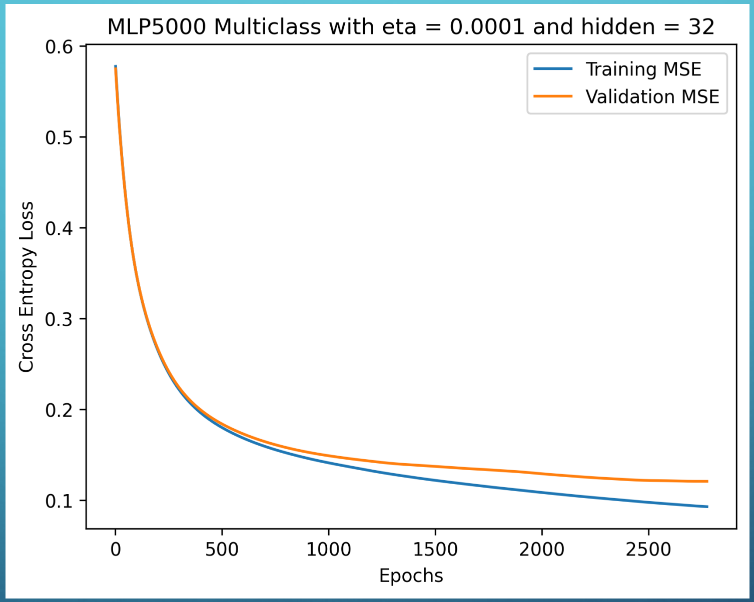


RESULTS-ECG5000 NORMAL VS ABNORMAL

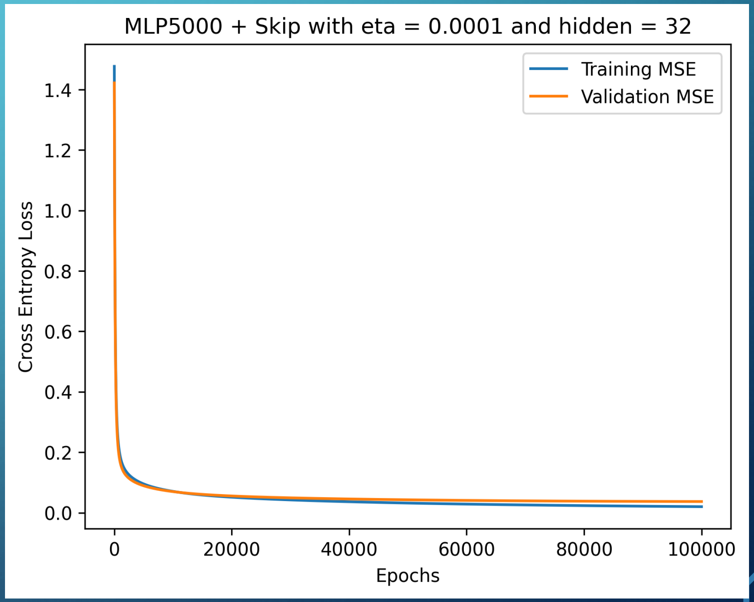
Shallow - Good



Deep – Better

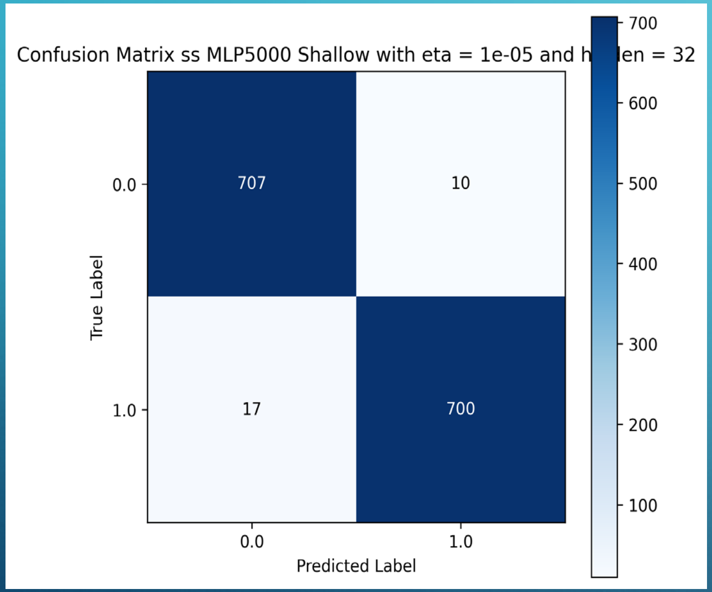


Skip Residuals - Best

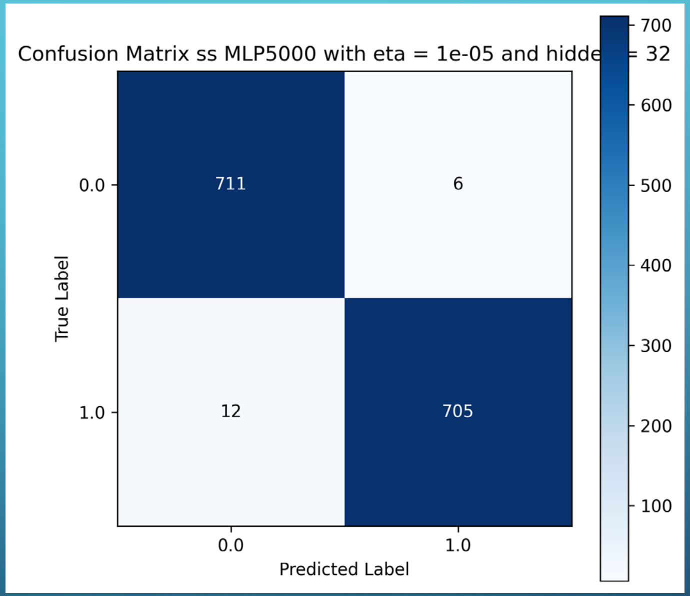


RESULTS-ECG5000 NORMAL VS ABNORMAL

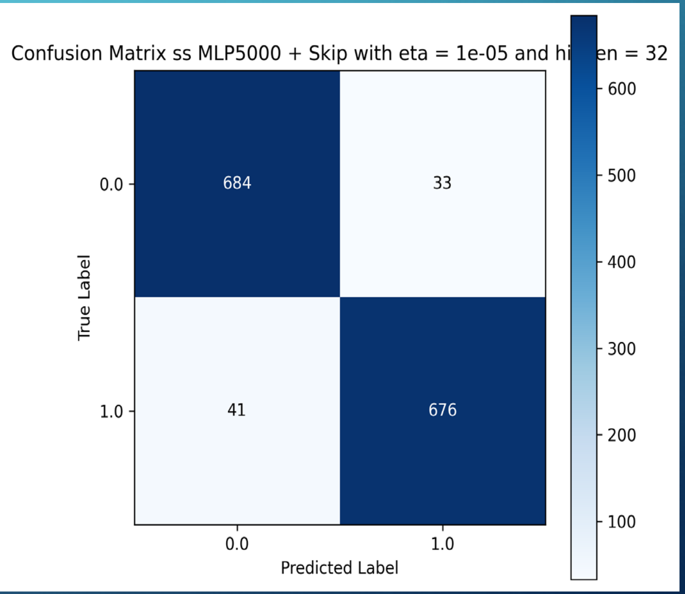
Shallow Accuracy = 98%



Deep Accuracy = 98%

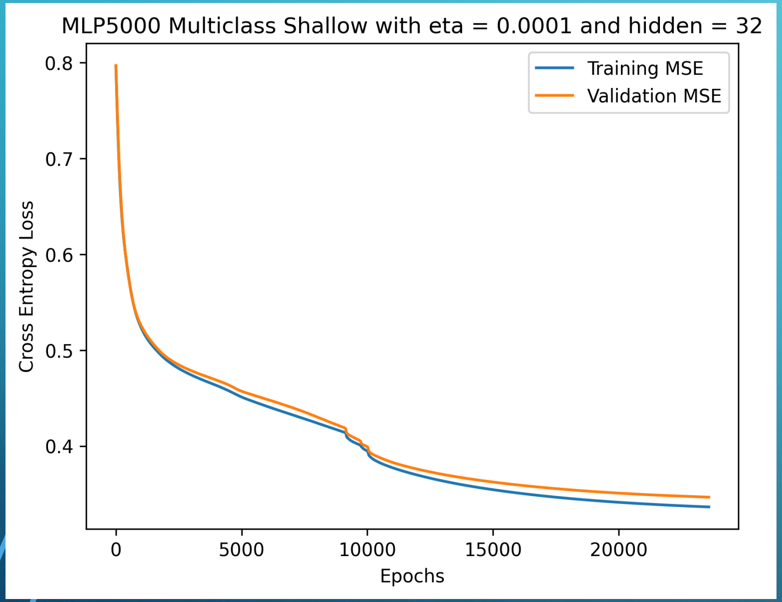


Deep with Skip Accuracy = 98%

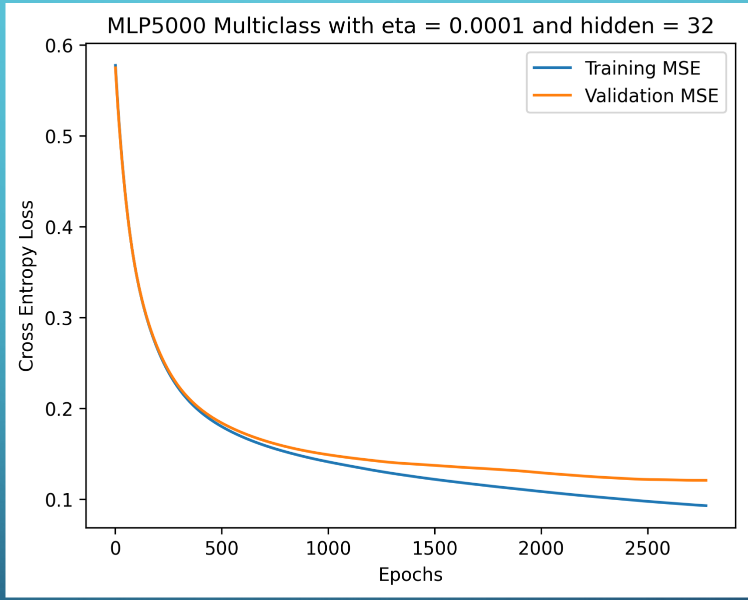


RESULTS-ECG5000 MULTICLASS

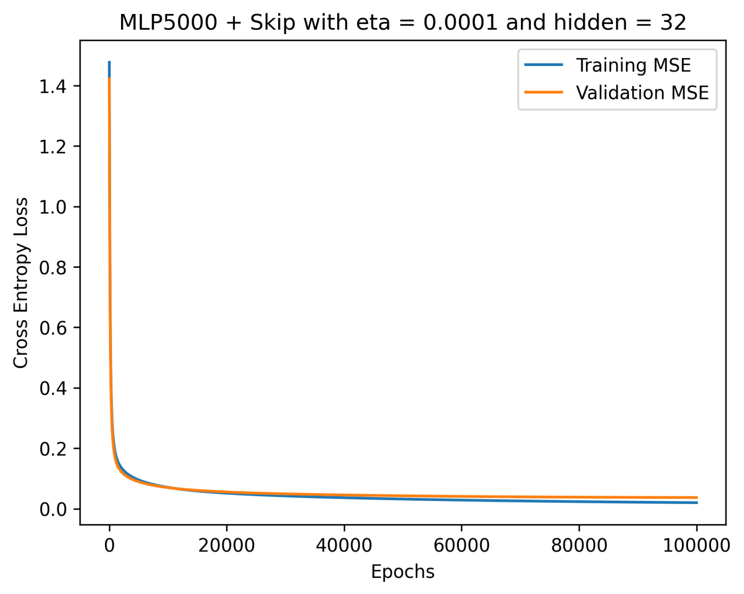
Shallow - Good



Deep – Better

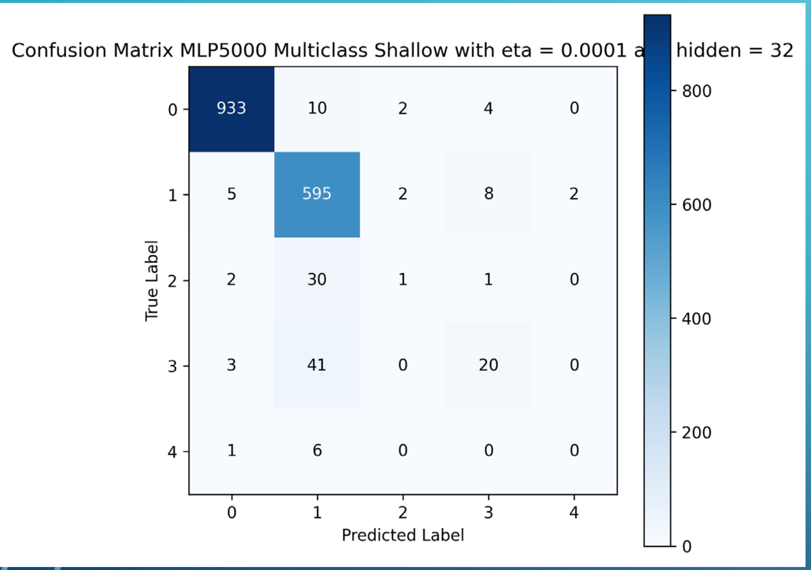


Skip Residuals - Best

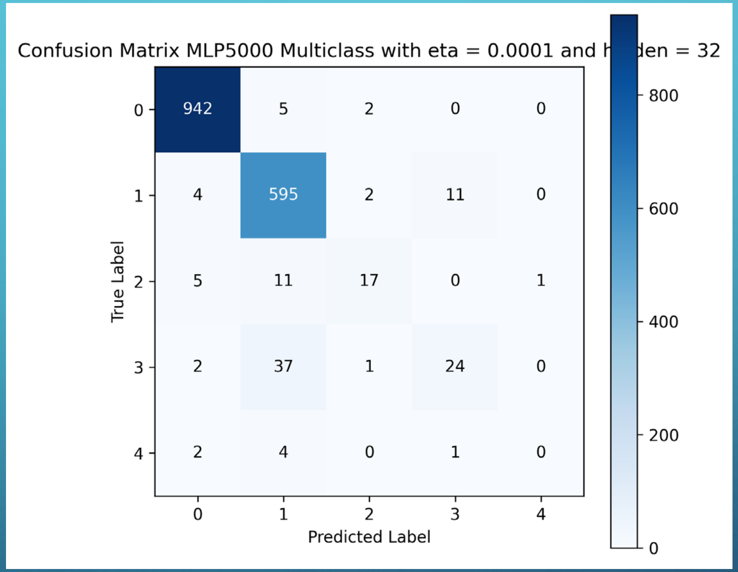


RESULTS-ECG5000 MULTICLASS

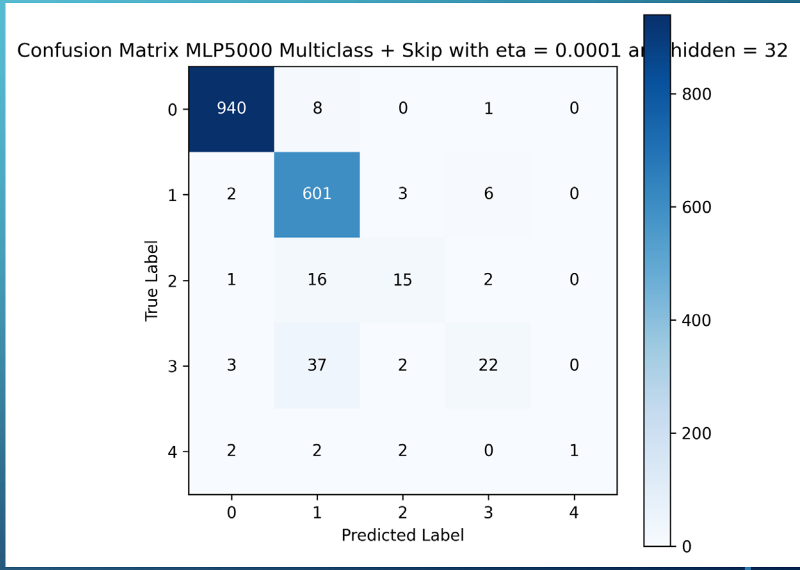
Shallow Accuracy = 92%



Deep Accuracy = 95%



Deep with Skip Accuracy = 95%

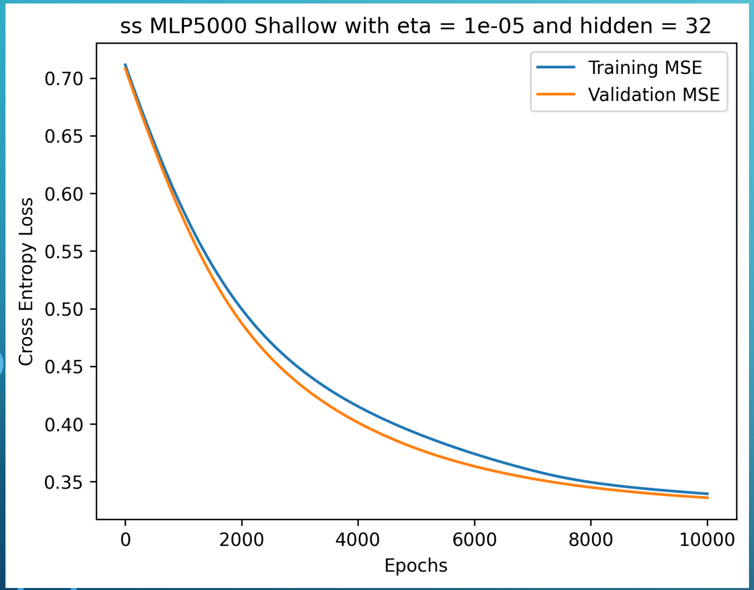


BUT THAT'S A LOT OF NORMAL CASES…

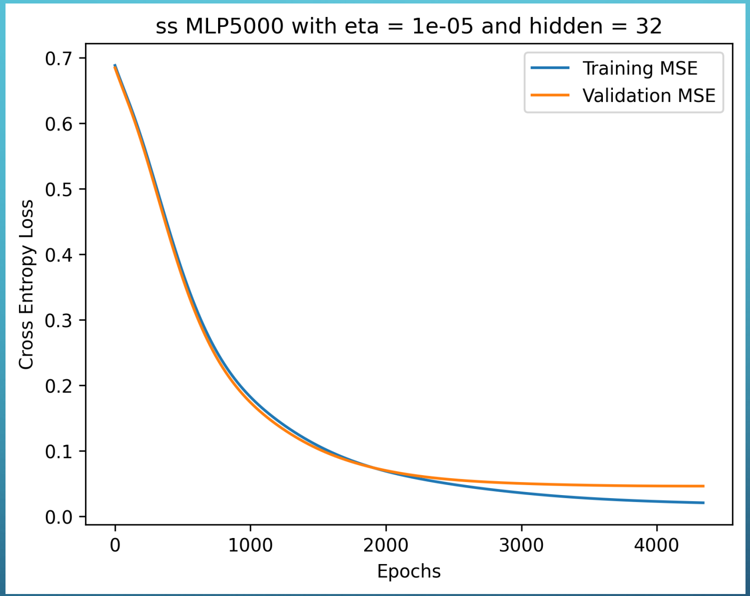
- Does it perform well just because of oversampled normal cases?

RESULTS-ECG5000 SUBSAMPLED NORMAL VS ABNORMAL

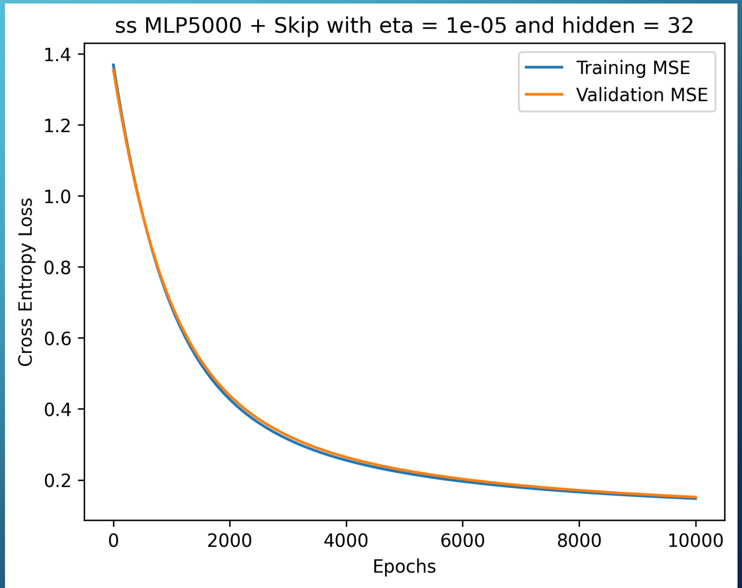
Shallow



Deep

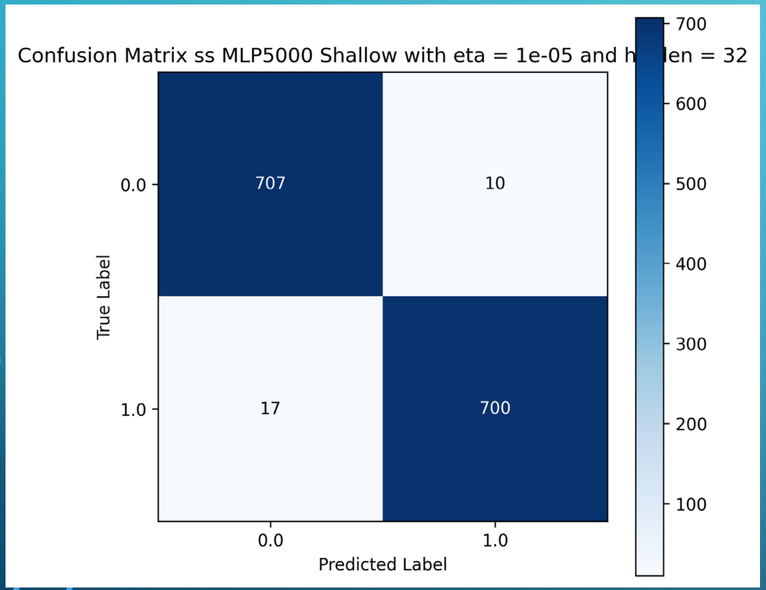


Deep with Skip

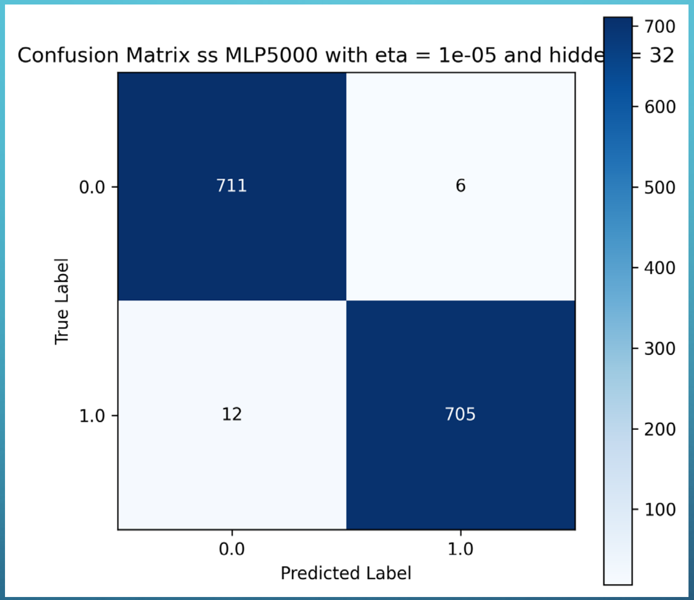


RESULTS-ECG5000 SUBSAMPLED NORMAL VS ABNORMAL

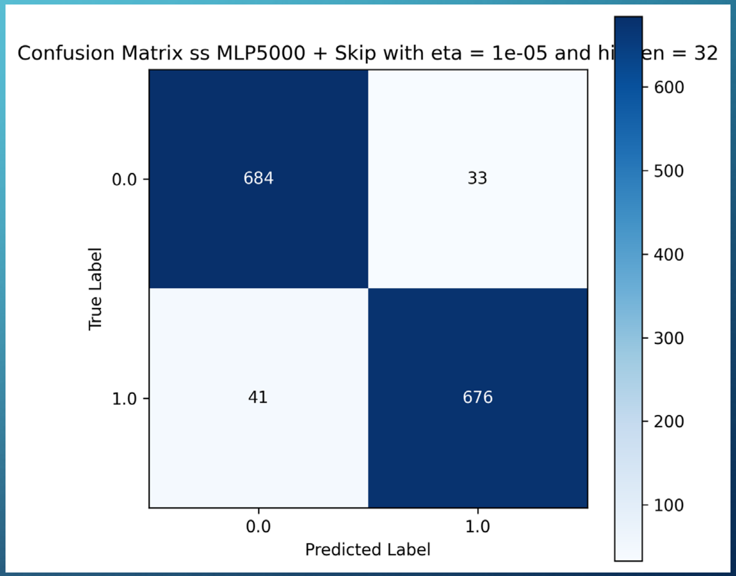
Shallow Accuracy = 98%



Deep Accuracy = 99%

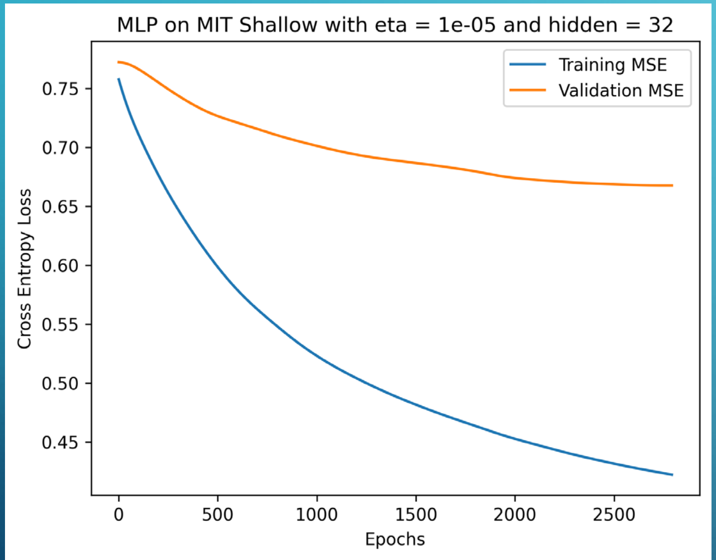


Deep with Skip Accuracy = 95%

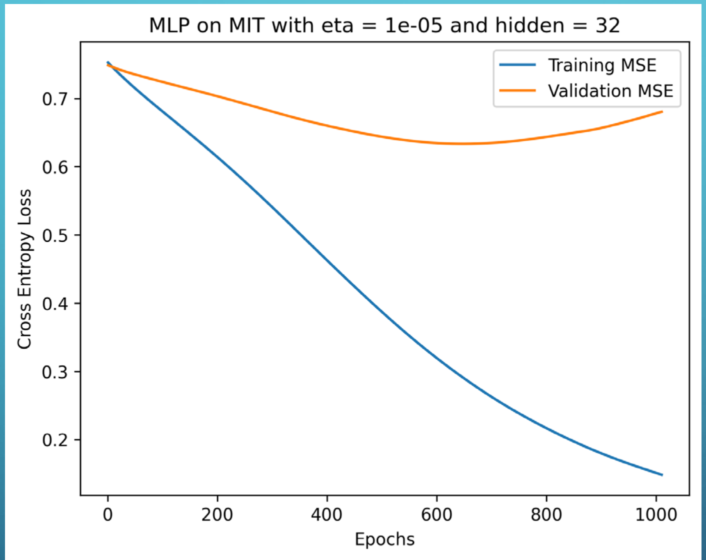


RESULTS MIT-BIH

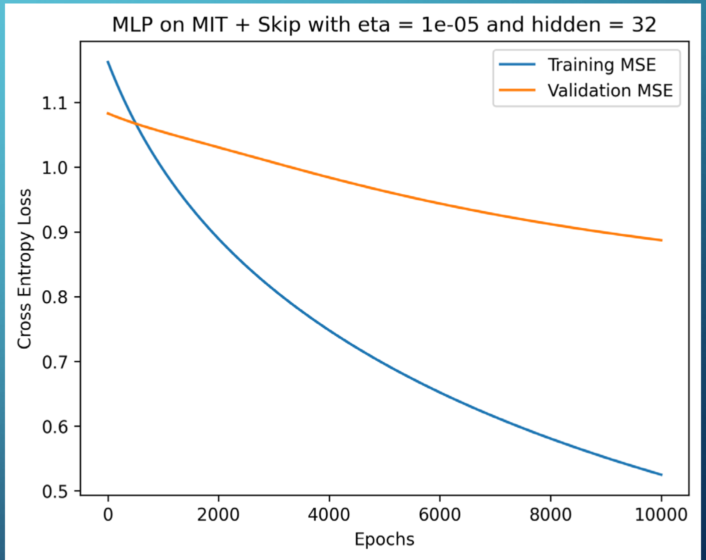
Shallow – Very overfit



Deep – Exploding Gradient



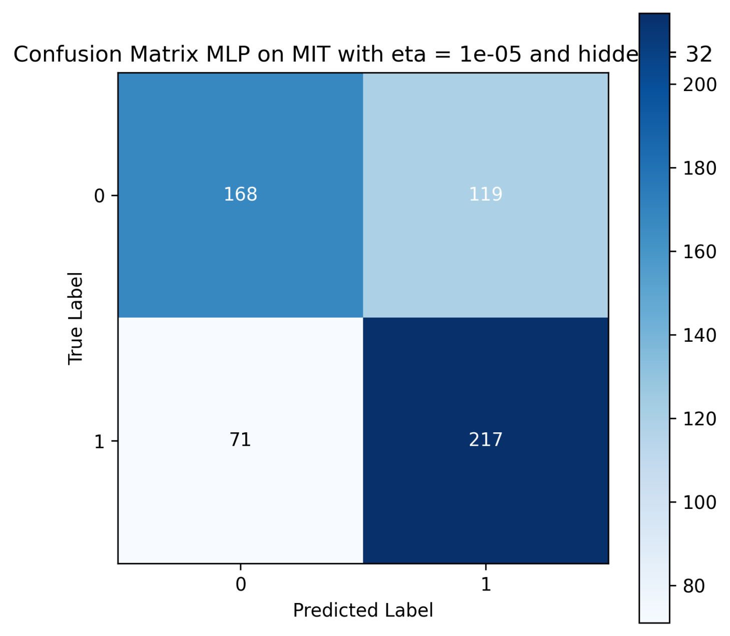
Skip Residuals… at least the line is pointing in the right direction



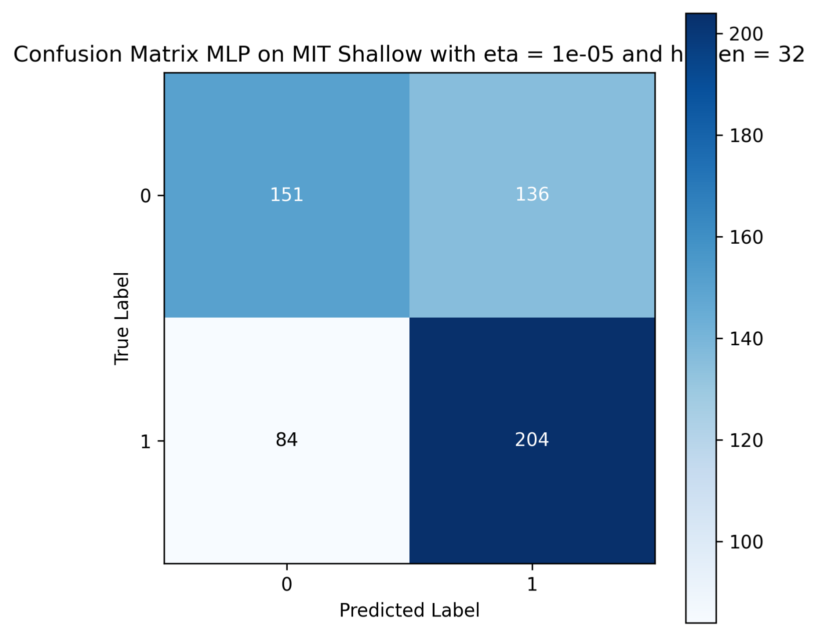
I RAN THIS FOR ANOTHER 10000 EPOCHS..
IT DOES NOT IMPROVE

RESULT MIT-BIH

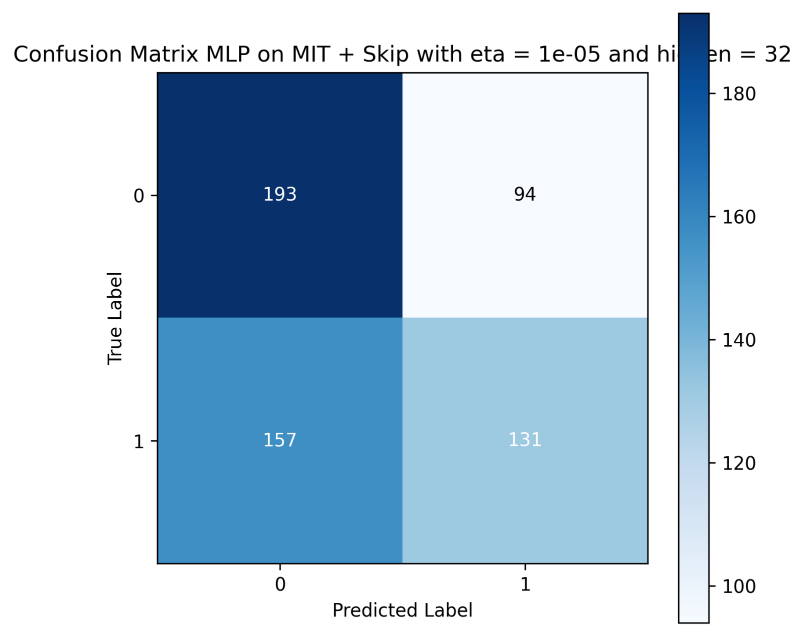
Shallow Accuracy = 67%



Deep Accuracy = 61%

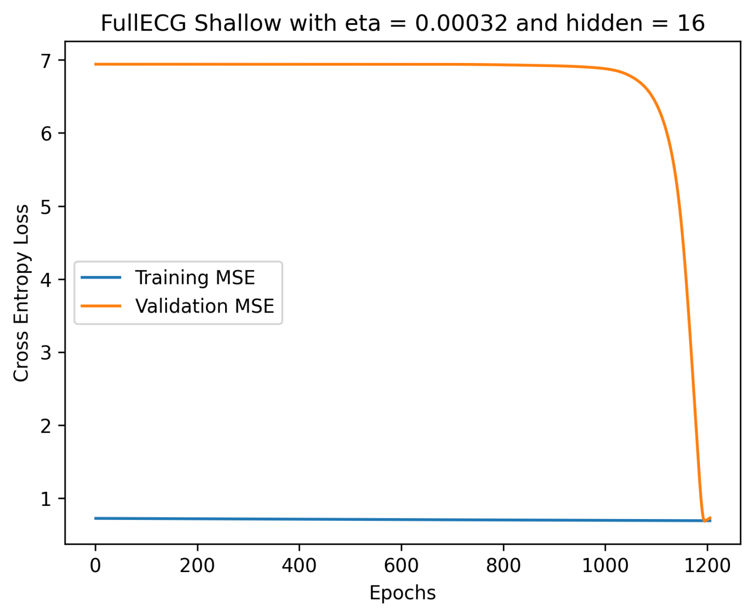


Deep with Skip Accuracy = 56%

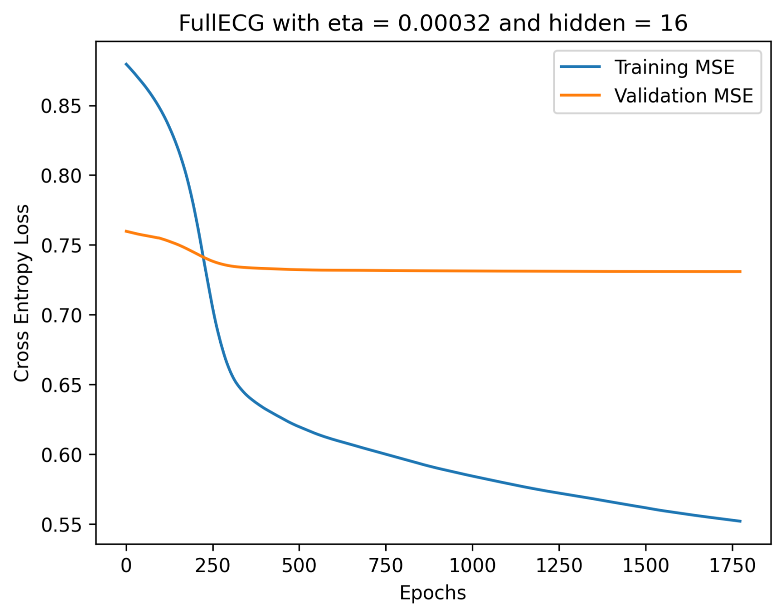


RESULT-FULL ECG

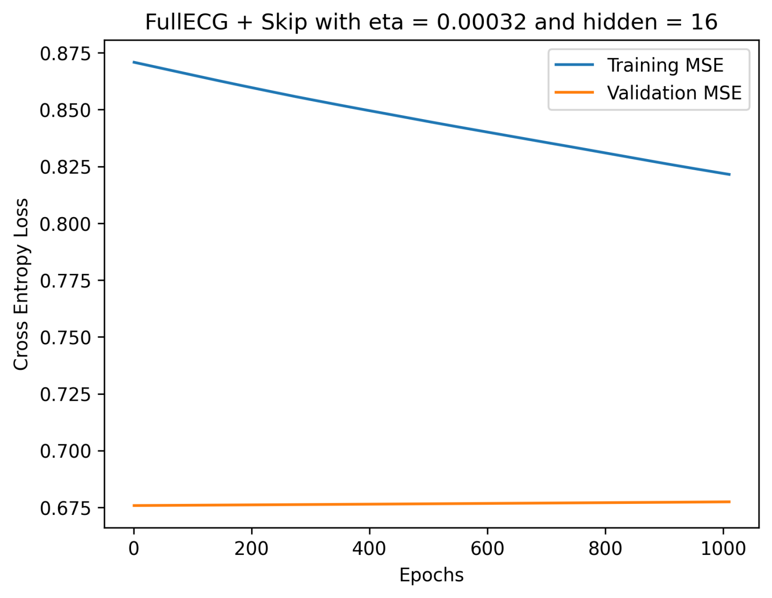
Shallow – Learned Something?



Deep – Overfit

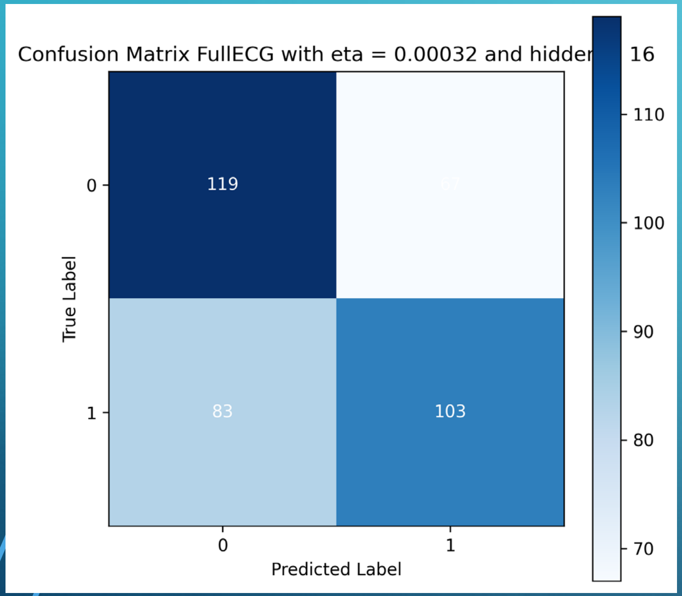


Deep with Skip - Guesses

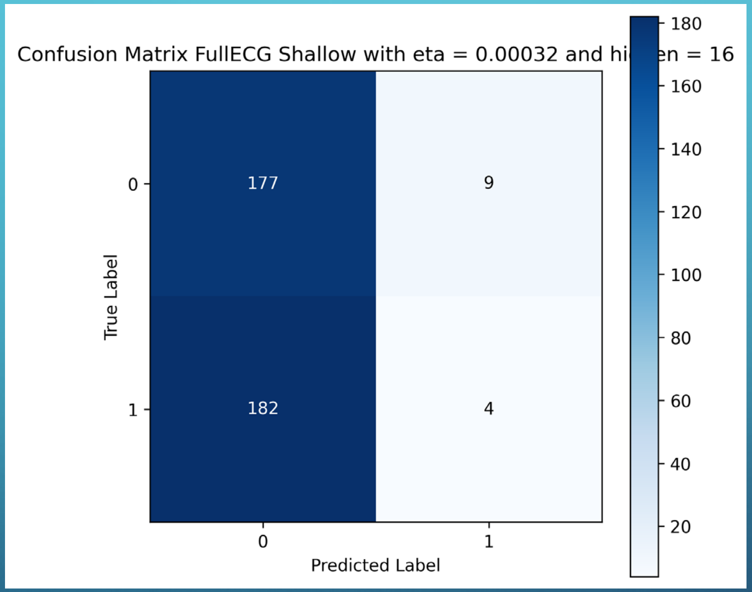


RESULTS-FULL ECG

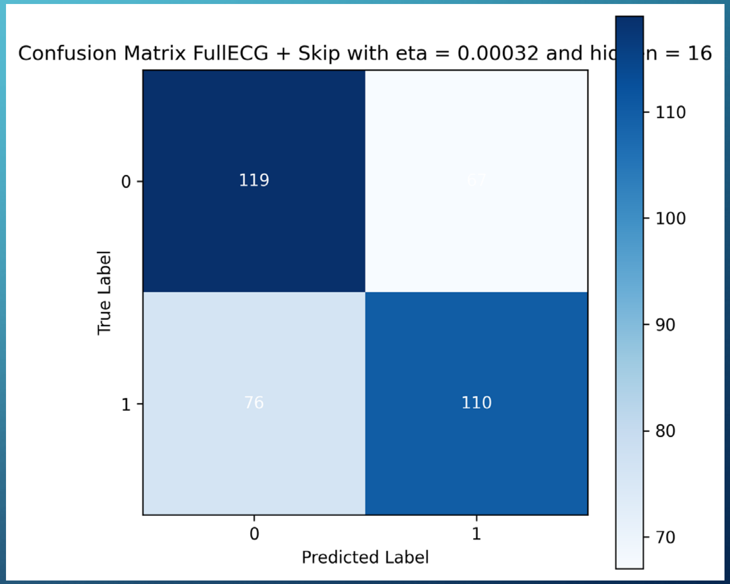
Shallow Accuracy = 72%



Deep Accuracy = 48%



Deep with Skip Accuracy = 75%



ANALYSIS – THE GOOD!

- Our MLP was effectively trained to identify normal vs ischemic single-cycle ECG tracing with an accuracy of 91%, made possible with the addition of skip residuals
- Our MLP was effective at identifying normal vs abnormal single-cycle ECG tracings. The shallow model performed with an accuracy of ***, leaving little motivation for more complex models or skip residuals

ANALYSIS – THE NOT SO GOOD…

- More complex datasets were not effectively analyzed by our MLP, despite numerous architecture modifications and optimizations, including a breadth search for optimal hyperparameters and attenuation of data missingness

DISCUSSION

- While MLP may be suitable for many classification problems, feature rich data may still prove too complex for this method.
- No globally-optimal model exists
- Multi-lead ECG data may be better classified using more complex models, such as Convolutional Neural Networks, Transformers, or Random Forest Models
- Future work is recommended to evaluate these datasets using more complex models

REFERENCES

Al-Zaiti, S. S., Martin-Gill, C., Zègre-Hemsey, J. K., Bouzid, Z., Faramand, Z., Alrawashdeh, M. O., Gregg, R. E., Helman, S., Riek, N. T., Kraevsky-Phillips, K., Clermont, G., Akcakaya, M., Sereika, S. M., Van Dam, P., Smith, S. W., Birnbaum, Y., Saba, S., Sejdic, E., & Callaway, C. W. (2023). Machine learning for ECG diagnosis and risk stratification of occlusion myocardial infarction. *Nature Medicine*, *29*(7), 1804–1813. https://doi.org/10.1038/s41591-023-02396-3

Dau, H. A., Keogh, E., Kamgar, K., Yeh, C.-C. M., Zhu, Y., Gharghabi, S., Ratanamahatana, C. A., et al. (2019). The UCR Time Series Classification Archive. University of California, Riverside. https://www.cs.ucr.edu/~eamonn/time_series_data_2018/

Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K., & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23), e215–e220. https://doi.org/10.1161/01.CIR.101.23.e215

Institute of Medicine (Ed.). (2000). *To err is human: Building a safer health system*. National Academies Press. http://www.nap.edu/catalog/9728

Institute of Medicine (Ed.). (2009). *Crossing the quality chasm: A new health system for the 21st century* (9. print). National Acad. Press.

Moody, G.B., & Mark, R.G. (2001). *The impact of the MIT-BIH Arrhythmia Database*. IEEE Engineering in Medicine and Biology Magazine, 20(3), 45–50. doi:10.1109/51.932724

Xiong, P., Lee, S. M.-Y., & Chan, G. (2022). Deep Learning for Detecting and Locating Myocardial Infarction by Electrocardiogram: A Literature Review. *Frontiers in Cardiovascular Medicine*, *9*, 860032. https://doi.org/10.3389/fcvm.2022.860032