American Sign Language Recognition Using Machine Learning Models

Department of Electronics, Telecommunications and Informatics

University of Aveiro, Portugal

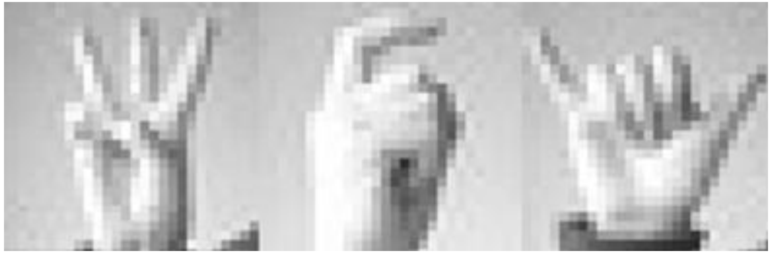
Solomonidi Eleftheria, 130364
Bilal el Hanane, 130152

Motivation & Problem Description

- Communication between Deaf ASL users and non-signers is often limited.
- Automated ASL recognition can support more accessible human–machine interaction.
- Objective: identify 24 ASL hand gestures from 28×28 grayscale images.
- Models trained and evaluated using the Sign Language MNIST dataset from Kaggle.

Dataset Overview & Image Characteristics

- The dataset contains grayscale images of static ASL hand signs, each 28×28 pixels.



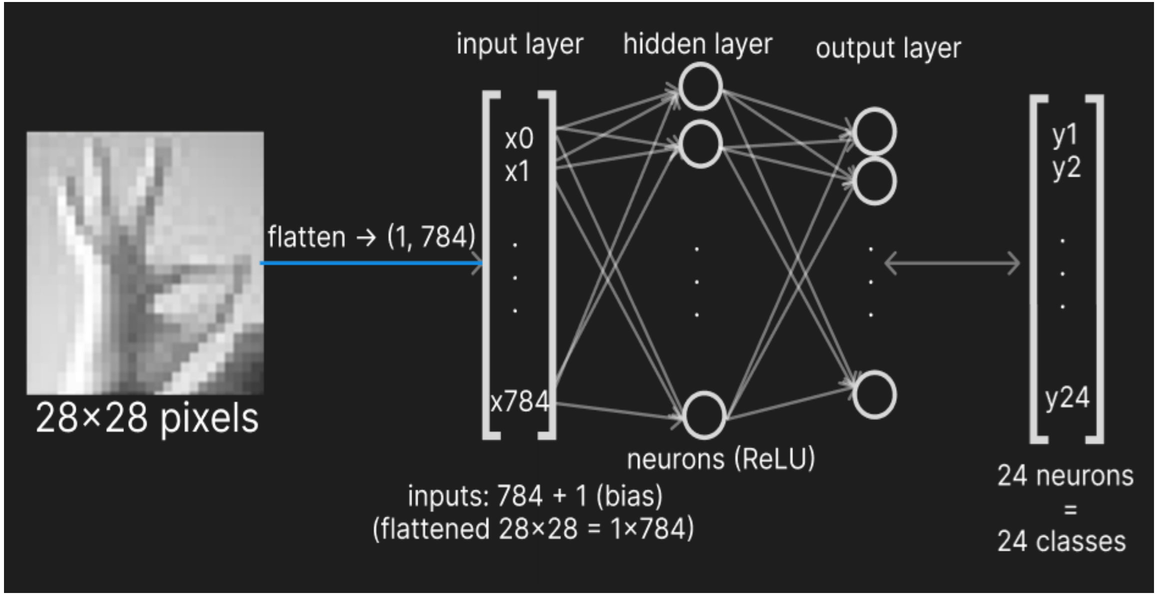
- Each class has hundreds of samples, allowing supervised model training.
- Data are normalized (vector (1,784)) and split into training, validation, and test sets before modeling.

Baseline & Transition to the MLP Model

- **Baseline:** Logistic Regression used as a simple one-vs-rest classifier. Provides a reference accuracy and confirms the task is learnable, but performance is limited.

- **Motivation for MLP:** Non-linear relationships in hand shapes require a deeper model. MLP is a neural network with hidden layers and non-linear activations that learns increasingly high feature representations, allowing it to model complex patterns that linear methods cannot.

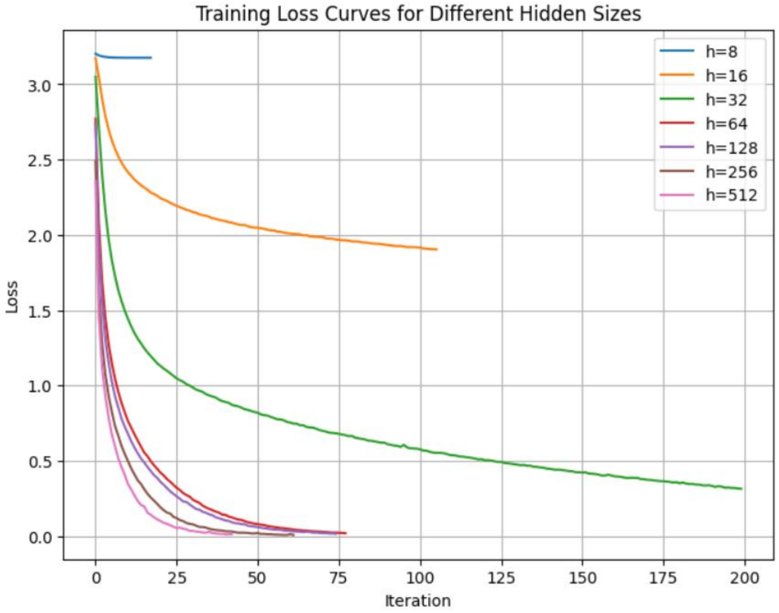
Neural Network Architecture



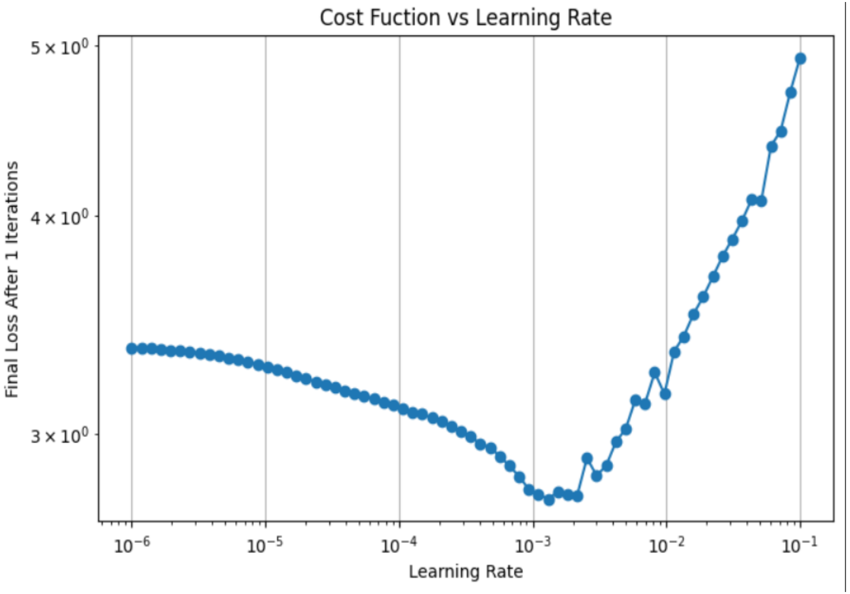
- Input images are flattened from 28×28 to 784 features.
- A hidden layer with ReLU activation learns non-linear patterns in hand shapes.
- Output layer has 24 neurons, one for each ASL class.

Hyperparameter tuning

- Hidden Layer Size



- Learning Rate (e.g. hidden layer size =64)

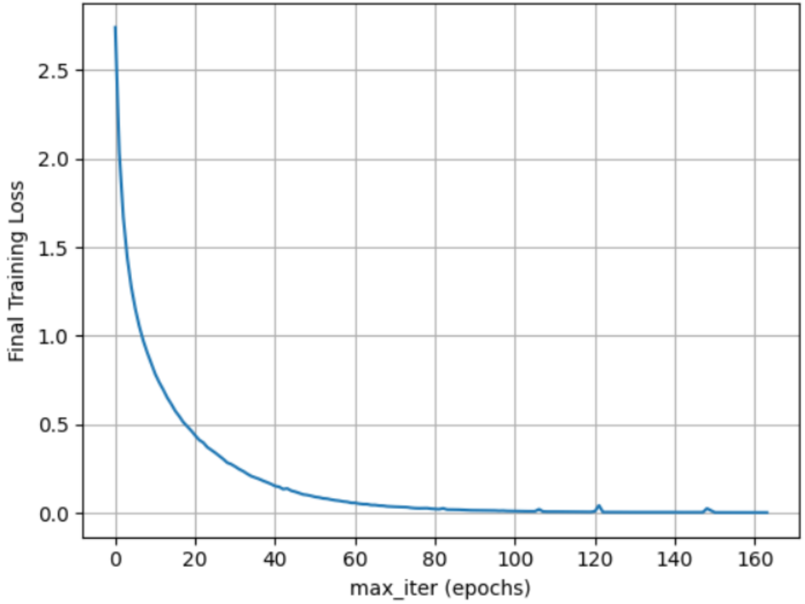


Hyperparameter tuning

- Regularization Strength (λ)

λ	val accuracy	val loss	iterations
0	0.99891	0.04663	91
0.00001	**0.99927**	**0.04217**	**91**
0.0001	0.99745	0.06699	80
0.001	0.99818	0.06443	80
0.01	0.99782	0.08300	105

- Maximum Number of Iterations



Final Model Configuration & Performance

Final Hyperparameters

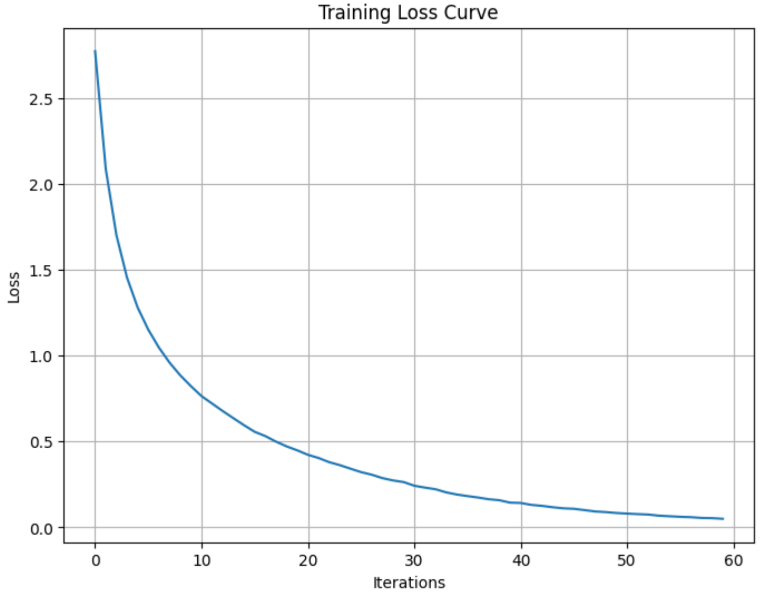
Hidden layer size	64
Learning rate	0.0001
Regularization (L2)	0.001
Maximum iterations	60

Training accuracy: 99.8867%

Validation accuracy: 99.6723%

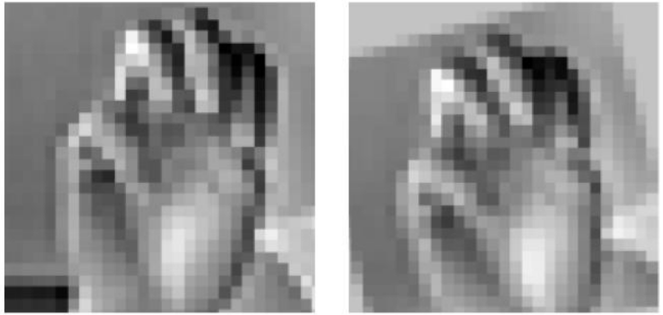
Test accuracy: 72.49%

Loss Function



Data Augmentation & Impact on Performance

- Double training set, 2*24709 images in total, using rotations, shifts, and brightness changes. These transformations create **slightly different versions** of the same hand sign.



- Helps the model handle small changes in **orientation, position, and lighting**.
- Test accuracy improved from **72.49%** → **82.77%**.

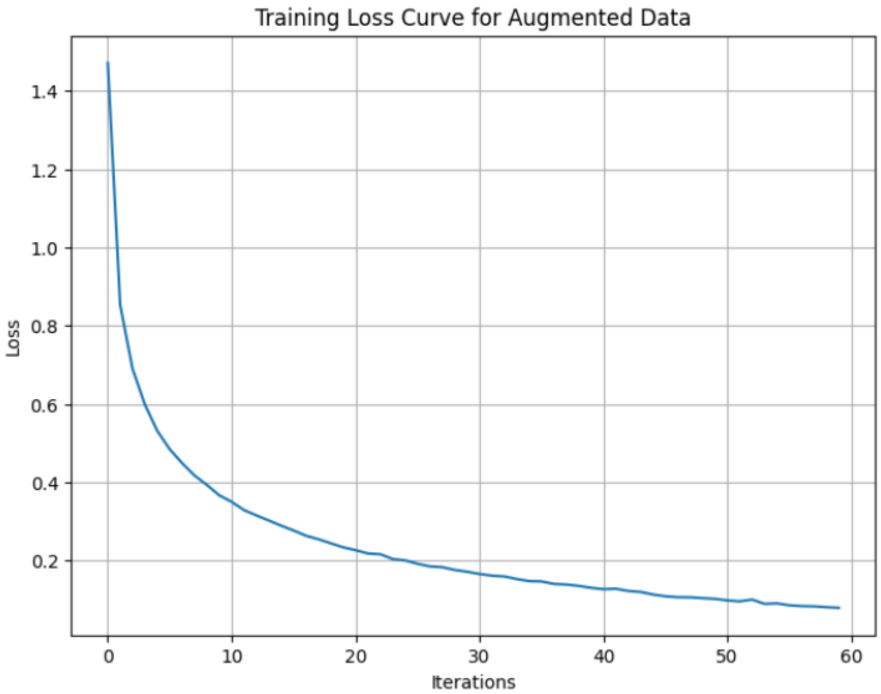
Configuration & Performance for Augmented Dataset

Final Hyperparameters

Hidden layer size	64
Learning rate	0.0001
Regularization (L2)	0.001
Maximum iterations	60

Training Accuracy: 97.9400%

- **Validation Accuracy: 99.8908%**
- **Test Accuracy: 82.7663%**



Test Set Inconsistencies

- Some test images do not match the official ASL gestures for their assigned labels. Example: class 24 shows a hand pose in the test set that is not an ASL letter.
- The model is evaluated on **incorrect or mismatched labels**.
- As a result, test accuracy **underestimates the true performance** of the model.

Letter Y

Train: 24

Test: 24

Conclusion & Future Work

- **Logistic Regression** provides a simple linear baseline, but cannot capture the complexity of ASL signs.
- The **MLP** is a better classifier for this task because it models non-linear gesture patterns.
- **Augmentation** significantly improved generalization on the test set.
- Future work: try **CNNs**, more augmentation, and improved dataset quality.