



Bangladesh University of Engineering and Technology

Department of Computer Science and Engineering

Academic Year 2022 - 2023

**CSE 472
-Machine Learning Sessional-**

Offline No. 3

**Feed Forward Neural Network with Dropout Layers and
Adam Optimization**

Name: Anwarul Bashir Shuaib

Roll: 1805010

Section: A1

Date of Submission: January 02, 2024

PERFORMANCE EVALUATION OF NEURAL NETWORK MODELS ON EMNIST DATASET

Abstract

This report presents a detailed evaluation of three distinct neural network models trained on the Extended Modified National Institute of Standards and Technology (EMNIST) dataset. Our objective was to identify the most effective model and combination for handwritten character recognition. Each model was trained using four different architectures, resulting in a total of twelve unique setups. The dataset comprised 124,800 training images, with 15% reserved for validation, and 20,800 images designated for independent testing. We carefully plotted training and validation cost curves, accuracy trajectories, and validation macro F1 score curves for each configuration to assess model performance and generalization capabilities. These visualizations provided insight on how the architecture and model architecture affected the training process. We experienced the overshooting phenomenon, as well as the effects of overfitting and underfitting. The outcome of this study was the selection of the best-performing model based on the validation metrics. This model was then evaluated on a separate test dataset. The results offer insightful conclusions about the effectiveness of different neural network architectures and when applied to the task of handwritten character recognition using the EMNIST dataset.

INSTRUCTIONS

- **Dataset:** EMNIST Balanced Dataset
- **Training Set Size:** 124,800
- **Test Set Size:** 20,800
- **Training Script:** `train_1805010.py`
- **Testing Script:** `test_1805010.py`
- **Saved Model:** `model_1805010.pickle`

Run the training script to train the model and save it to a file. Then run the testing script to evaluate the model on the test dataset. The model will be loaded from the saved file and the test results will be printed to the console.

For each of the model, we trained with learning rates of 0.005, 0.001, 0.0005, and 0.0001. The best model was selected based on the validation macro F1 score. The best model was then evaluated on the test dataset. The results are presented in the following sections.

PERFORMANCE EVALUATION OF THE FIRST MODEL

Best validation macro F1 score: 0.91412, achieved at epoch 30 with learning rate 0.0005, batch size 1024, and dropout rate 0.2.

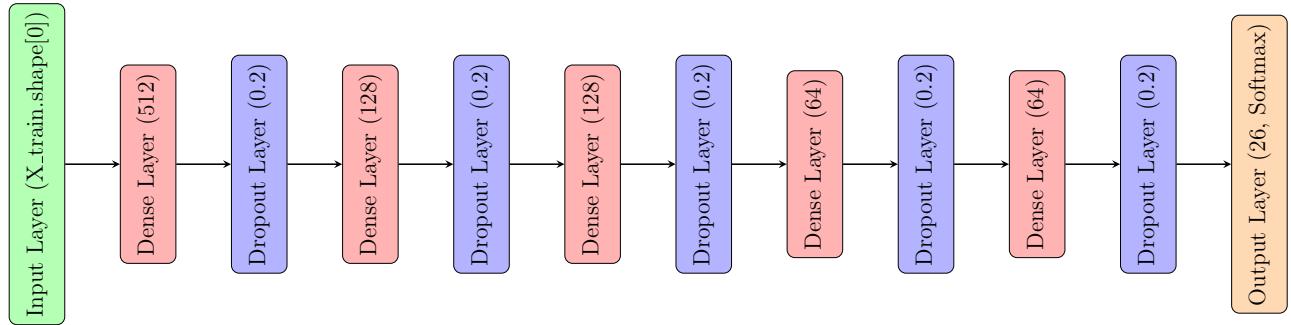


Figure 1: Architecture of the first model

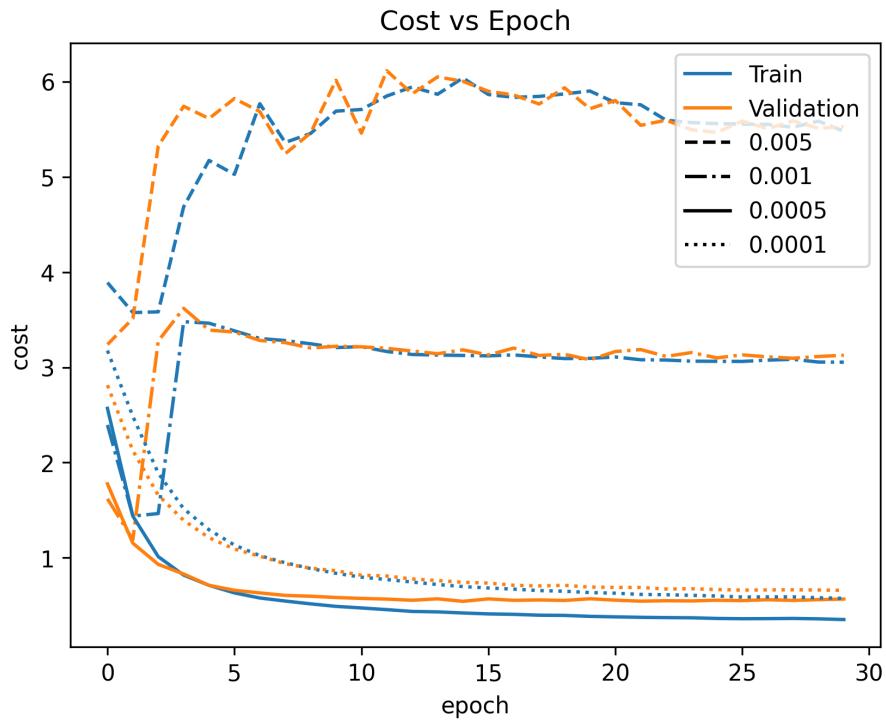


Figure 2: Training and validation loss curves for the first model

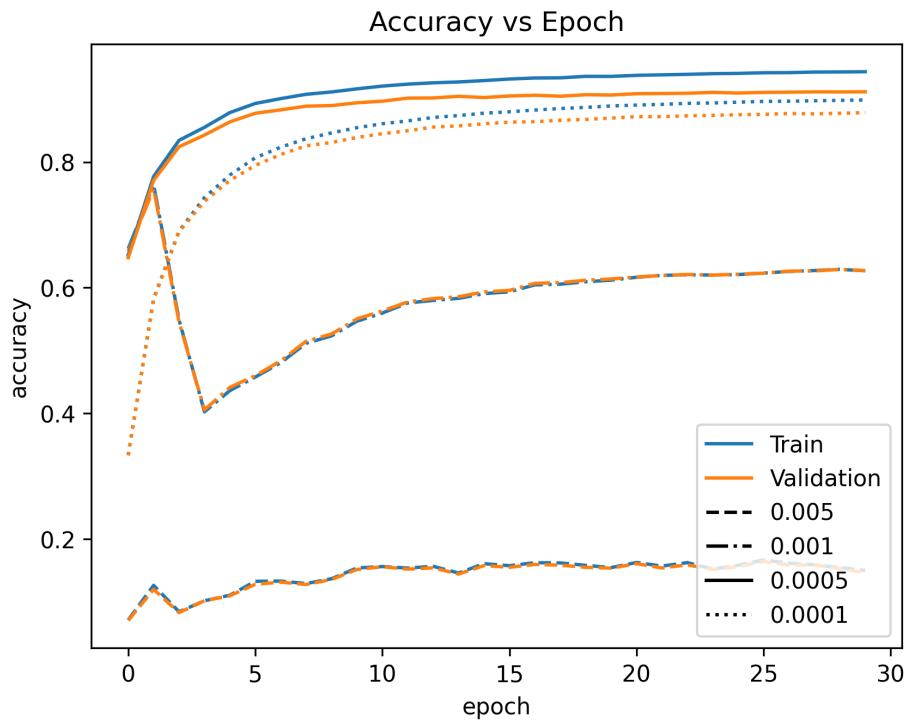


Figure 3: Training and validation accuracy curves for the first model

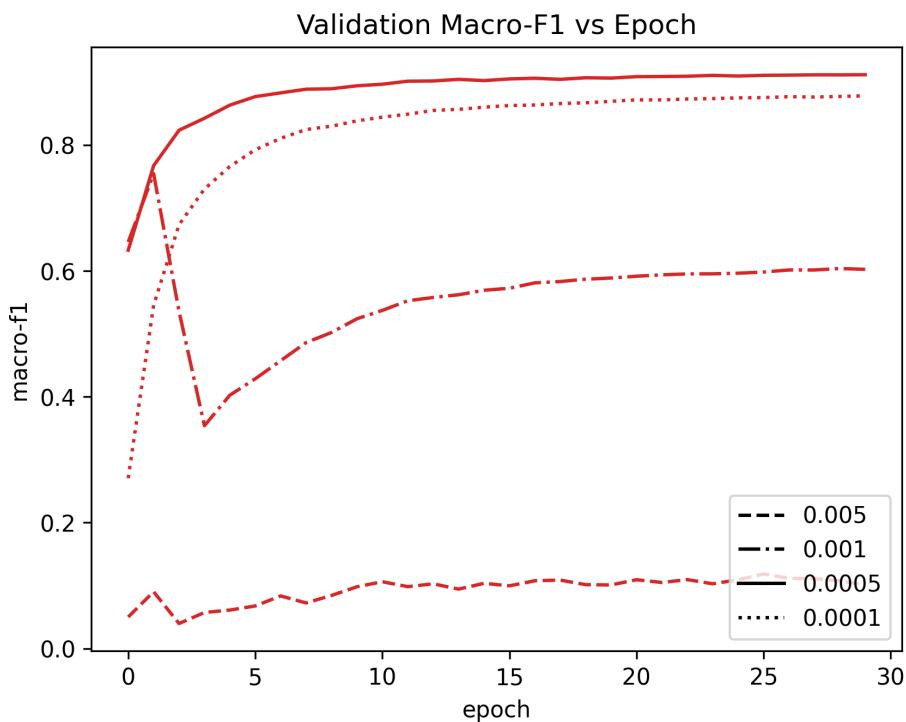


Figure 4: Validation macro F1 score curves for the first model

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	
True	22	34	0	0	1	4	0	0	1	0	3	17	0	0	0	0	0	9	1	703	0	0	0	0	5		
a	41	109	0	0	5	2	0	6	0	2	0	32	0	0	0	1	0	0	17	2	407	0	0	1	0	10	
b	7	2	0	0	3	0	0	0	2	3	0	0	0	0	0	0	0	3	2	778	0	0	0	0	0		
c	13	25	0	0	0	10	0	2	1	0	1	12	0	0	0	0	0	2	7	1	710	1	0	1	0	5	
d	46	33	0	0	4	6	0	1	2	0	5	20	0	0	0	0	0	1	9	1	663	0	0	0	0	11	
e	41	10	0	0	0	58	0	2	1	1	18	11	0	0	0	16	0	5	13	23	56	0	0	3	0	17	
f	31	31	0	0	2	56	0	0	0	0	1	9	0	0	0	8	0	2	66	4	491	0	0	0	0	39	
g	54	84	0	0	1	69	0	34	4	8	5	11	0	0	0	8	0	2	57	13	543	1	0	2	0	46	
h	9	0	0	0	12	0	28	6	11	0	10	0	0	0	10	0	0	48	36	20	0	0	10	1	1		
i	8	10	0	0	3	40	0	33	7	15	1	10	0	0	0	18	0	1	107	13	112	2	0	6	0	17	
j	5	20	0	0	5	44	0	7	5	5	28	16	0	0	0	6	0	9	38	8	139	2	0	3	0	41	
k	10	22	0	0	0	13	0	6	5	2	1	27	0	0	0	4	0	0	16	32	146	0	0	10	3	5	
l	2	2	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	1	0	793	0	0	0	0	1		
m	7	0	0	1	6	0	1	0	1	2	6	0	0	0	0	0	2	1	1	765	0	0	0	0	1		
n	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	795	0	0	0	0	0		
o	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	795	0	0	0	0	0		
p	23	1	0	0	1	66	0	0	2	16	4	0	0	0	11	0	2	19	26	0	0	2	0	7			
q	44	27	0	0	3	138	0	8	0	7	8	8	0	0	0	23	0	0	57	6	394	0	0	6	0	61	
r	47	11	0	0	2	66	0	6	1	3	23	13	0	0	0	12	0	1	14	18	127	21	0	8	1	26	
s	26	38	0	0	3	6	0	2	0	3	0	20	0	0	0	4	0	0	148	1	249	0	0	0	0	3	
t	23	29	0	0	1	79	0	5	17	7	15	13	0	0	0	57	0	1	34	89	80	17	0	24	8	16	
u	2	2	0	0	0	19	0	0	0	0	5	0	0	0	0	0	4	0	0	766	1	0	0	0	1		
v	2	6	0	0	0	6	0	0	2	0	10	4	0	0	0	10	0	0	53	3	7	178	435	0	1	11	2
w	0	0	0	0	0	4	0	0	0	0	3	0	0	0	0	0	1	0	0	792	0	0	0	0	0		
x	13	13	0	0	0	0	17	0	29	2	12	0	15	0	0	0	118	0	0	17	350	31	13	0	41	6	12
y	11	27	0	0	0	67	0	1	38	1	0	35	0	0	0	22	0	1	21	81	92	65	0	95	117	7	
z	12	103	0	0	0	12	0	13	0	38	0	38	0	9	0	0	6	0	0	80	3	163	0	0	6	0	19

Confusion matrix for $lr = 0.005$

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
True	576	50	13	28	17	10	12	37	1	1	3	0	12	4	16	8	0	0	1	0	0	0	1	0	0	7
a	89	664	1	18	0	5	3	2	0	0	2	2	1	2	0	1	2	0	0	0	0	0	1	0	0	7
b	19	3	7	2	15	2	2	1	1	0	4	0	0	1	4	3	0	6	0	0	1	0	0	0	0	3
c	40	599	3	599	1	1	1	5	1	1	1	1	0	2	5	15	7	2	0	1	0	3	4	2	0	0
d	3	49	127	2	599	6	0	0	2	1	3	0	2	1	2	2	1	2	1	0	0	0	0	0	0	2
e	2	9	3	3	20	693	6	3	8	4	1	1	1	0	0	33	0	5	2	4	0	0	0	0	0	
f	112	149	27	4	5	34	133	2	1	8	1	0	2	0	2	9	0	0	5	1	1	1	2	0	0	
g	210	232	1	40	1	5	0	7	0	7	0	15	2	0	2	0	0	0	4	1	1	2	0	0	0	
h	0	3	4	20	11	2	1	1	1	1	1	1	0	1	0	1	0	5	7	1	1	7	1	17	0	
i	0	14	8	65	0	4	3	0	31	660	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	
j	4	24	9	6	17	6	3	660	3	0	449	0	0	2	1	0	3	1	1	0	0	0	3	0	0	
k	1	13	52	61	0	5	1	8	52	5	44	58	0	0	0	10	0	2	0	1	0	0	3	1	13	
l	9	5	0	0	0	3	0	3	42	0	0	1	0	0	2	1	3	1	2	1	0	0	0	0	0	
m	30	9	0	105	0	0	2	768	0	0	15	0	120	233	1	2	1	3	0	0	10	5	1	1	2	
n	35	21	36	3	1	0	1	0	1	0	0	10	11	664	1	0	0	0	0	0	1	0	0	0	0	0
o	5	0	7	5	69	8	2	0	0	0	0	0	1	8	0	664	0	1	0	0	0	0	0	0	0	
p	289	31	11	1	11	49	37	1	4	0	0	0	0	1	3	18	1	0	1	3	1	0	2	4	0	
q	32	9	14	3	72	36	1	50	8	2	51	1	8	23	0	60	0	1	0	1	0	0	6	0	2	
r	8	60	16	34	26	89	42	39	167	72	0	4	11	4	1	62	2	87	1	53	0	0	0	5	9	
s	0	2	4	55	1	0	2	13	1	7	0	5	25	0	1	0	0	0	0	0	4	2	1	0	0	
t	1	0	1	0	8	0	5	28	707	0	4	0	0	0	0	2	0	6	21	1	0	0	0	2		
u	1	8	3	0	2	4	0	23	1	7	1	3	0	1	4	0	5	1	3	1	10	5	1	1		
v	0	0	0	3	0	0	1	1	0	3	0	1	7	0	1	30	75	2	2	0	0	1	4	0	0	
w	0	0	0	3	0	0	1	2	0	4	19	0	0	0	0	1	9	3	75	0	1	0	0	4		
x	1	3	0	1	3	0	1	2	2	0	1	0	1	3	0	76	18	4	0	0	0	0	0	0		
y	4	1	0	2	0	1	0	2	2	0	0	2	0	1	0	3	1	78	18	18	4	0	0	0		
z	1	1	2	7	0	5	1	2	1	0	1	0	0	0	1	1	4	1	61	7	17	0	0	2		

Confusion matrix for $lr = 0.001$

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
True	4	3	7	7	1	4	6	0	0	0	5	5	5	17	2	0	0	3	0	1	1	0	0	4	5	

PERFORMANCE EVALUATION OF THE SECOND MODEL

Best validation macro F1 score: 0.90124, achieved at epoch 30 with learning rate 0.001, batch size 1024, and dropout rate 0.2.

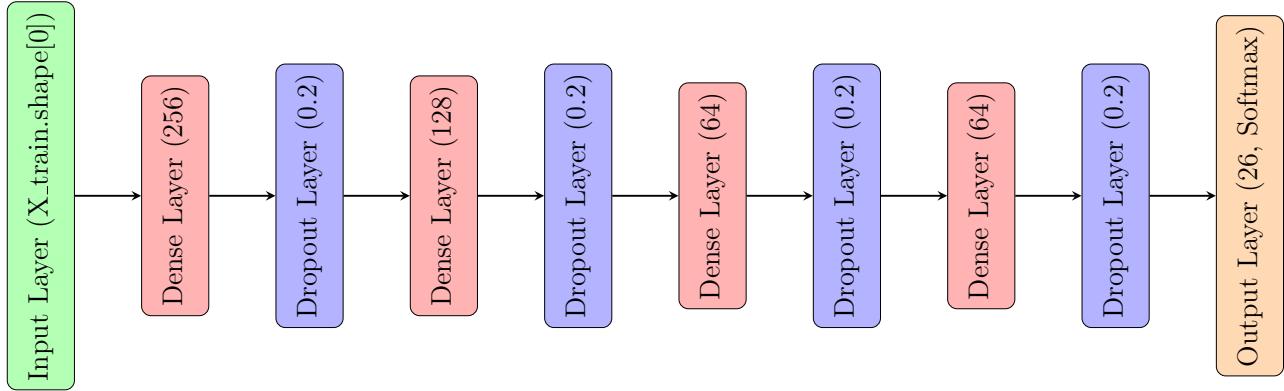


Figure 5: Architecture of the secondfirst model

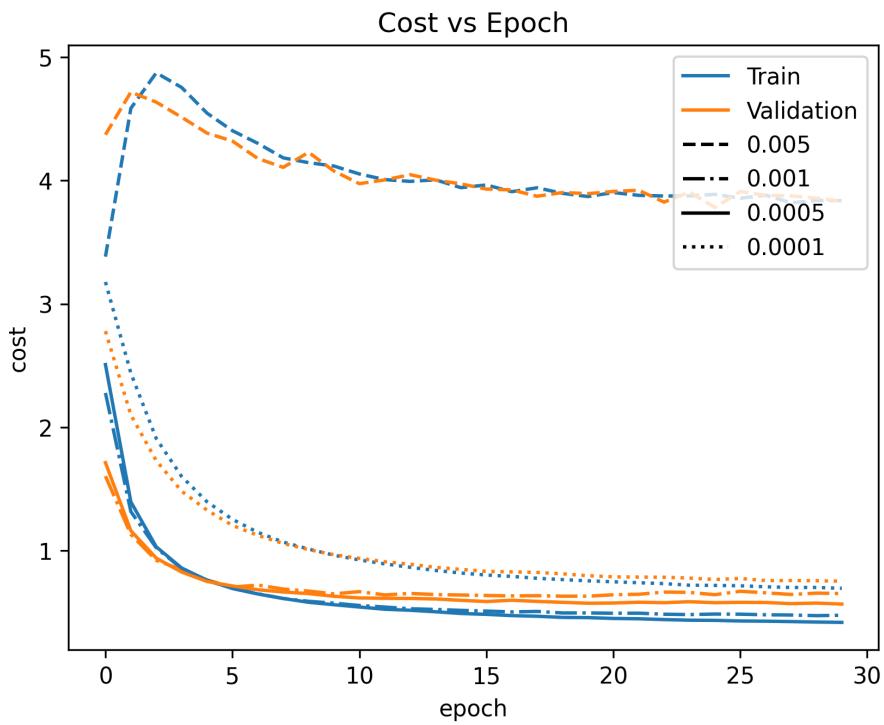


Figure 6: Training and validation loss curves for the second model

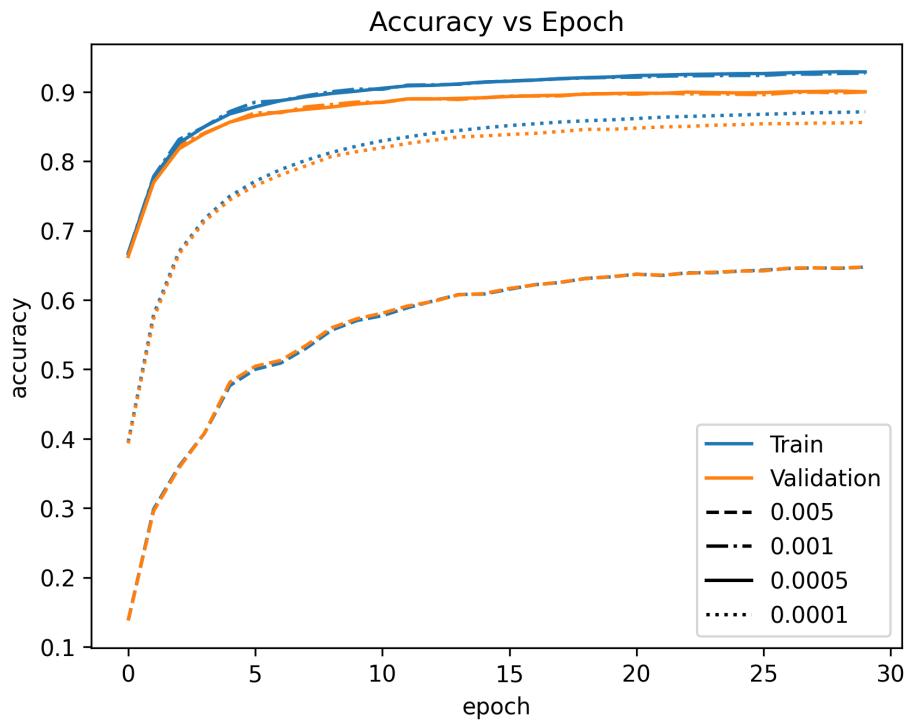


Figure 7: Training and validation accuracy curves for the second model

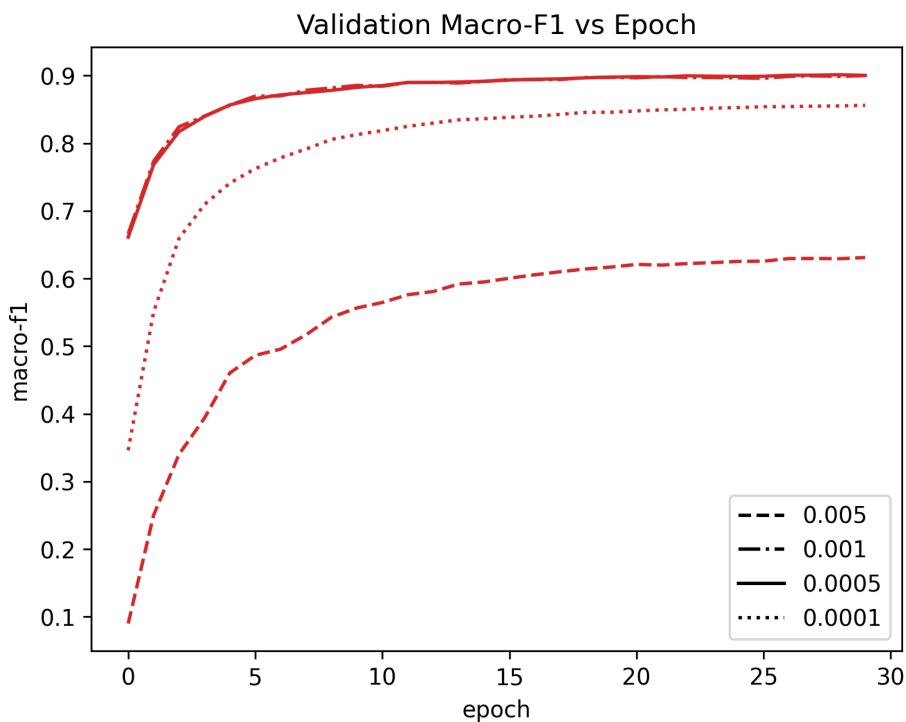


Figure 8: Validation macro F1 score curves for the second model

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	
True	a	14	61	6	20	25	16	28	69	1	2	6	0	13	36	27	6	9	1	0	1	11	0	3	0	0	5
True	b	1	21	5	3	7	11	7	3	0	0	2	2	3	3	0	2	1	0	0	0	0	0	4			
True	c	16	2	728	2	29	3	1	0	1	0	1	0	0	1	4	2	0	3	0	0	5	0	0	0	0	1
True	d	23	338	1	55	2	6	7	13	1	5	3	1	2	2	35	6	1	0	0	0	13	1	2	0	0	1
True	e	43	18	549	0	541	9	5	0	4	1	1	1	2	3	2	1	6	6	0	2	0	1	0	4		
True	f	10	27	3	1	25	613	2	4	8	4	0	2	1	2	0	37	3	18	1	2	1	5	0	0	1	0
True	g	62	150	19	7	16	68	30	3	4	12	0	0	3	8	2	4	22	0	11	7	9	0	2	0	4	1
True	h	65	163	0	21	1	7	1	280	4	0	11	3	12	10	0	2	5	2	0	1	1	1	5	4	0	1
True	i	6	2	0	2	6	29	2	0	707	25	1	0	0	0	0	0	1	6	0	3	0	0	3	1	6	
True	j	12	7	4	0	10	43	4	57	59	0	1	1	0	0	0	0	1	2	6	3	0	0	1	0	2	
True	k	14	10	4	4	27	2	0	113	6	0	97	1	5	0	0	1	0	3	0	0	5	1	6	1	0	
True	l	9	21	16	16	50	36	1	5	5	2	14	72	0	0	0	1	0	1	0	1	5	0	0	1	7	2
True	m	11	2	0	0	0	1	7	0	0	2	0	767	3	0	1	2	0	0	0	2	0	1	1	0	0	
True	n	30	0	16	10	7	6	28	1	0	26	0	69	320	5	1	8	2	0	0	32	6	20	2	4	0	
True	o	11	13	9	23	1	4	4	0	1	0	0	0	2	6	72	1	1	0	0	0	3	0	0	0	0	
True	p	3	1	0	6	2	3	1	1	0	0	0	1	5	2	694	2	6	0	1	0	1	0	1	0		
True	q	5	39	7	11	36	171	26	2	7	11	2	0	2	4	16	11	56	4	2	6	13	0	3	4	7	0
True	r	42	5	20	0	109	24	3	17	9	0	57	0	4	38	1	15	0	5	13	1	13	0	5	0	9	7
True	s	9	12	1	2	3	11	19	0	9	58	0	0	0	3	4	0	0	2	666	0	1	0	0	0	0	
True	t	17	3	3	70	67	31	24	33	52	7	6	12	13	2	9	19	100	0	189	0	3	2	2	16	0	
True	u	64	8	1	98	1	0	1	4	0	16	26	4	7	10	12	0	8	0	1	0	572	13	7	0	0	
True	v	4	1	0	2	0	3	1	8	1	5	0	1	2	0	1	0	13	0	5	91	652	0	2	8	0	
True	w	5	7	0	9	0	0	1	21	0	1	5	0	14	11	0	0	0	0	0	58	7	63	1	2	0	
True	x	12	1	0	4	2	2	3	15	16	3	110	0	3	1	0	0	0	1	1	2	2	14	0	0	0	
True	y	12	3	0	9	1	2	58	4	29	26	1	2	5	4	0	5	12	6	3	13	4	97	1	41	862	
True	z	10	7	6	13	23	0	5	3	25	16	4	4	3	0	2	1	0	0	4	2	0	0	12	0		

Confusion matrix for $lr = 0.005$

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	
True	a	71	4	2	7	8	1	6	12	0	0	1	0	1	3	10	2	21	1	0	2	4	0	1	0	0	4
True	b	9	1	43	0	10	3	1	8	11	1	0	2	4	0	1	3	0	0	1	0	0	0	0	1	0	
True	c	2	1	77	2	29	1	1	0	0	0	1	4	0	0	1	3	2	1	2	0	1	1	0	0	0	
True	d	10	8	0	22	0	1	0	1	4	0	1	1	3	38	0	0	1	0	0	4	0	3	1	0	2	
True	e	6	2	15	0	57	2	2	0	0	0	2	0	0	5	2	0	0	2	1	0	0	3	0	1	0	
True	f	0	1	0	1	6	723	2	0	4	0	0	0	0	0	0	23	2	6	3	24	0	0	0	1	3	
True	g	19	24	7	2	4	4	0	1	2	1	5	0	1	0	2	1	1	8	0	0	0	0	0	3	1	
True	h	5	7	0	3	1	0	0	0	0	0	0	0	0	0	4	13	7	19	0	0	0	2	0	1	3	0
True	i	0	0	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
True	j	0	3	0	10	1	0	1	3	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	4	
True	k	3	5	1	0	2	0	0	0	17	0	1	1	1	11	4	2	0	2	7	2	7	3	1	1	12	1
True	l	1	0	0	8	0	0	1	0	5	16	0	0	0	0	0	0	0	1	0	2	0	1	0	0	0	
True	m	2	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	1	2	0	2	1	
True	n	12	0	0	4	0	0	0	0	15	0	0	4	0	0	0	21	723	1	0	0	7	0	0	2	2	
True	o	3	1	1	12	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	1	0	
True	p	0	1	0	8	1	0	0	0	0	0	0	0	0	0	0	0	3	1	16	8	4	0	2	0	1	
True	q	23	3	3	4	5	3	36	0	1	2	0	2	0	1	10	0	3	69	2	2	1	2	0	0	1	2
True	r	11	1	2	0	6	5	1	1	1	0	4	1	0	1	4	1	1	11	1	7	0	6	6	6	6	
True	s	2	2	0	0	1	2	1	2	8	0	0	0	0	0	1	2	0	1	0	0	0	0	0	0	1	
True	t	2	3	1	0	5	3	1	2	2	0	0	0	0	0	0	20	1	1	0	0	0	0	0	0	1	
True	u	0	2	0	15	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
True	v	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
True	w	0	0	1	1	0	0	0	2	0	1	0	0	1	15	4	74	0	0	0	13	8	15	0	0	0	
True	x	3	0	1	2	0	1	0	1	1	1	1	1	15	3	0	2	0	0	1	2	1	0	5	0	19	
True	y	0	3	0	1	0	0	5	1	8	0	2	0	0	0	1	0	8	0	14	2	3	743	1	1		
True	z	5	1	0	4	8	0	4	1	1	0	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	

Confusion matrix for $lr = 0.001$

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
True	a	6	4	8	9	2	9	15	0	1	2	0	5	10	18	5	34	2	2	2	5	0	2	3	0	8
True	b	13	708	1	13	4	0	13	18	2	2	1	4	1	4	3	1	3	1	4	1					

PERFORMANCE EVALUATION OF THE THIRD MODEL

Best validation macro F1 score: 0.90037, achieved at epoch 30 with learning rate 0.0001, batch size 1024, and no dropout.

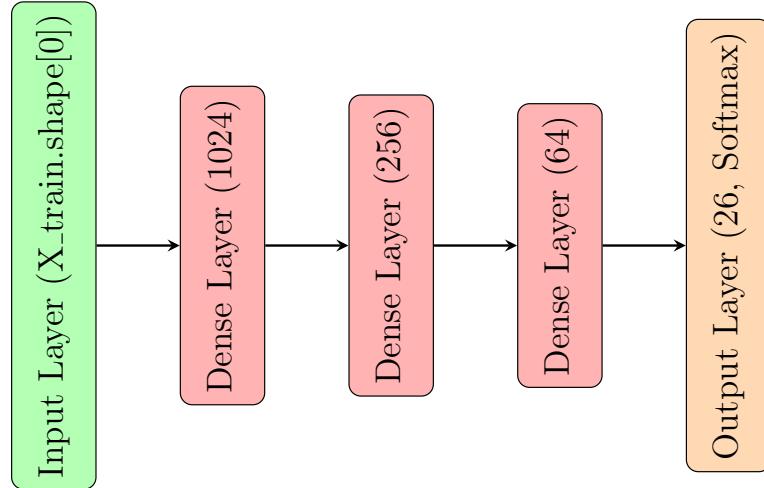


Figure 9: Architecture of the third model

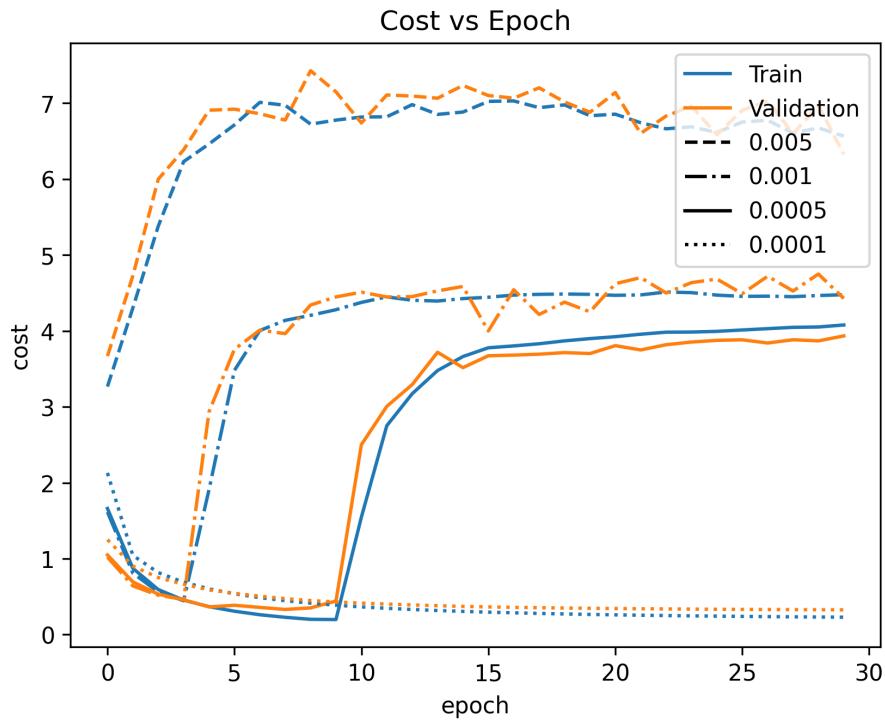


Figure 10: Training and validation loss curves for the third model

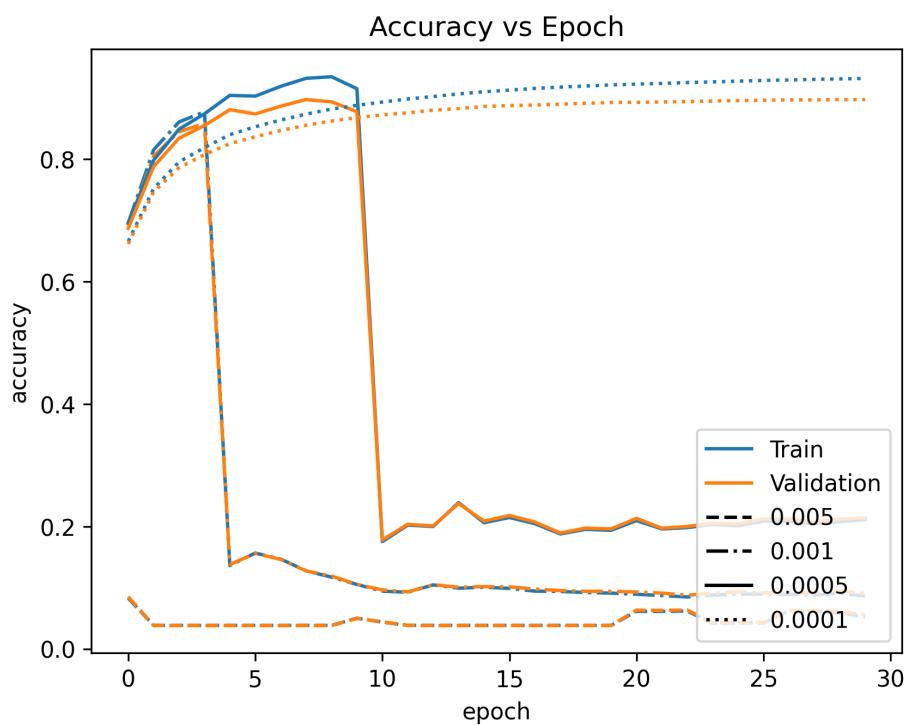


Figure 11: Training and validation accuracy curves for the third model

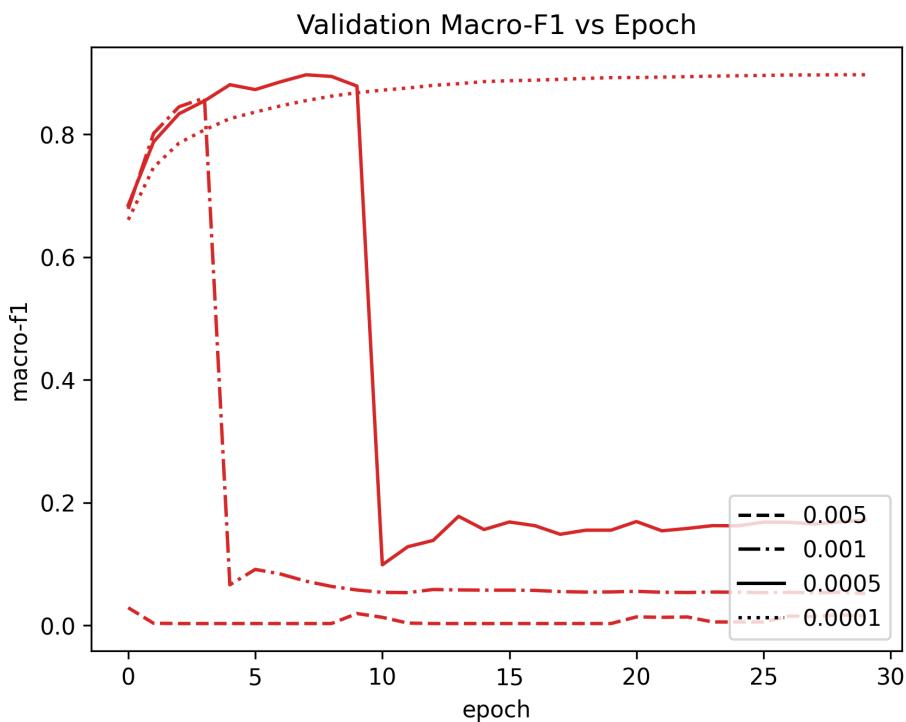
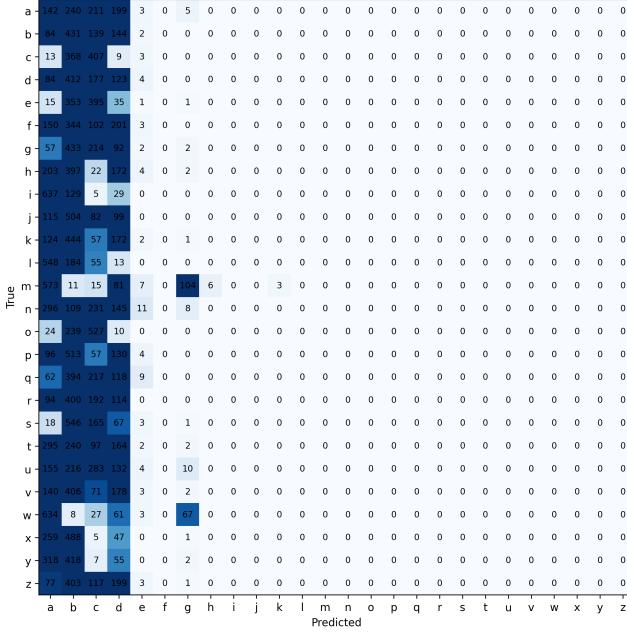
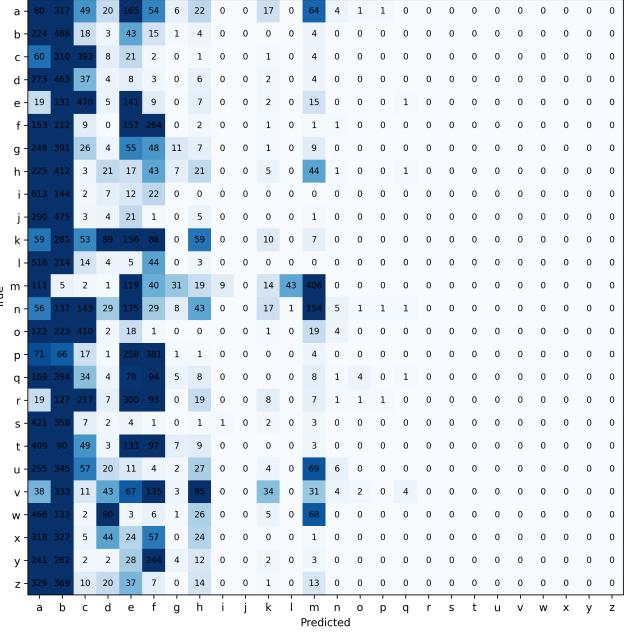


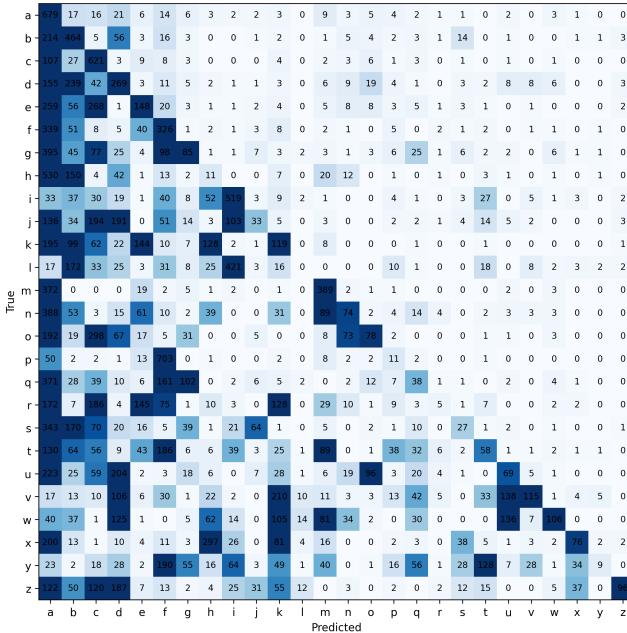
Figure 12: Validation macro F1 score curves for the third model



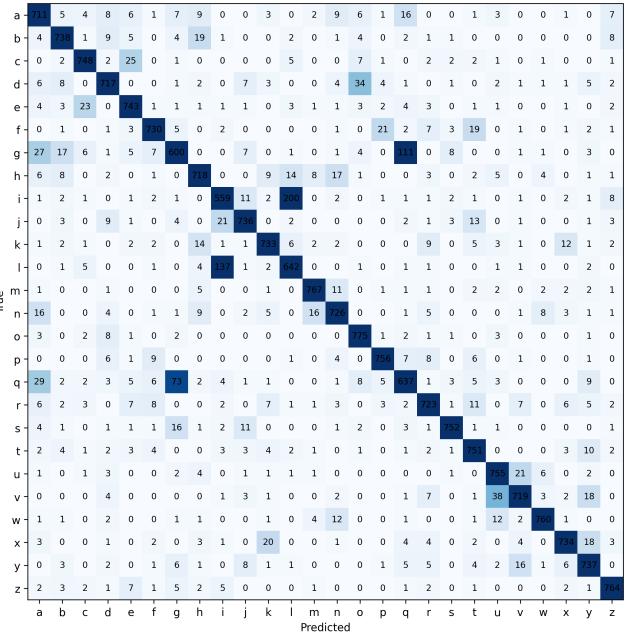
Confusion matrix for $lr = 0.005$



Confusion matrix for $lr = 0.001$



Confusion matrix for $lr = 0.0005$



Confusion matrix for $lr = 0.0001$

PERFORMANCE EVALUATION OF THE BEST MODEL

Best validation macro F1 score: 0.91412, achieved at epoch 30 with learning rate 0.0005, batch size 1024, and dropout rate 0.2.

Model	Learning Rate	Batch Size	Dropout Rate	Validation Macro F1 Score
1	0.0005	1024	0.2	0.91412
2	0.001	1024	0.2	0.90124
3	0.0001	1024	0.0	0.90037

Table 1: Best validation macro F1 scores for each model over all learning rates

The test macro F1 score for the best model was 0.91139 and accuracy was 0.91135. The confusion matrix for the test dataset is shown below.

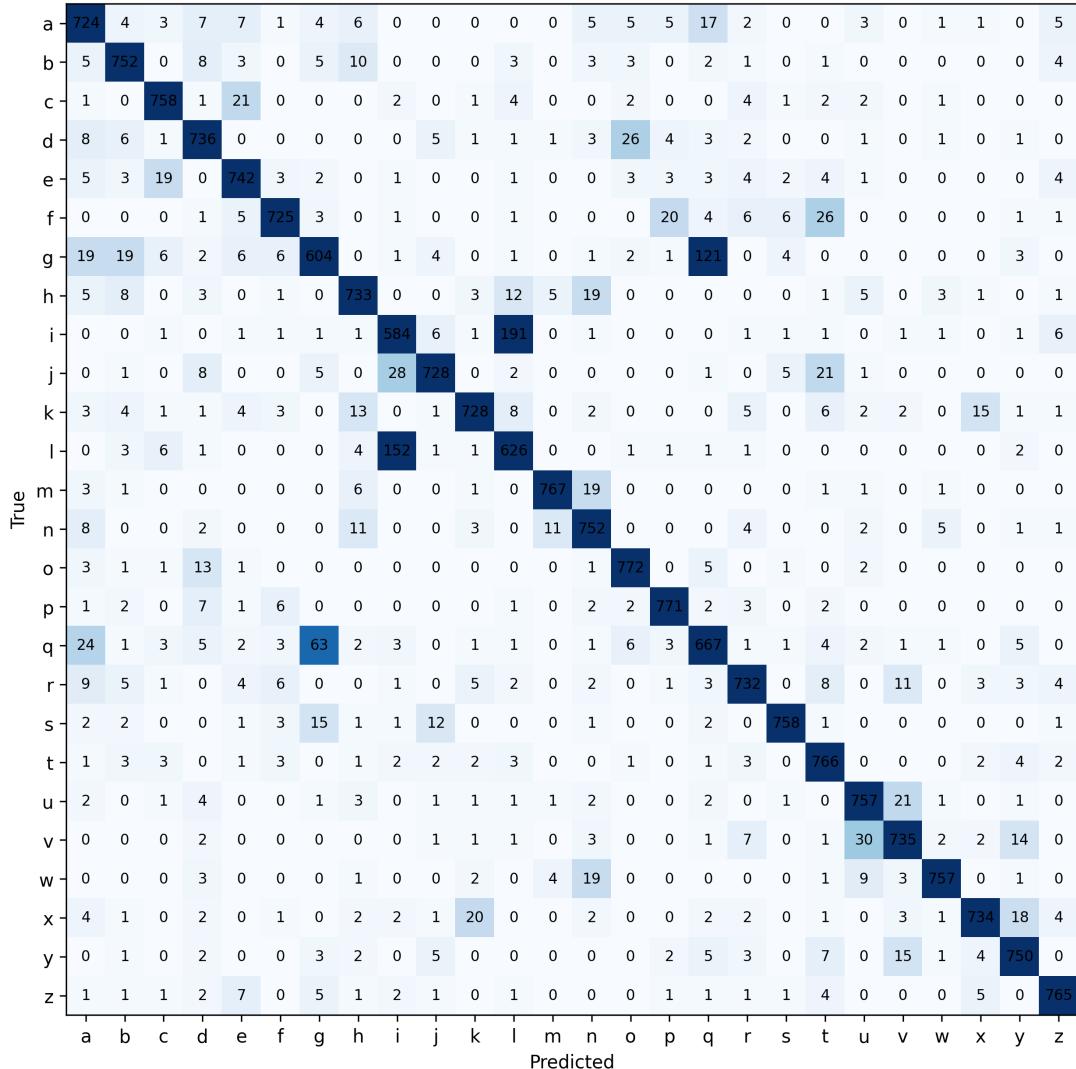
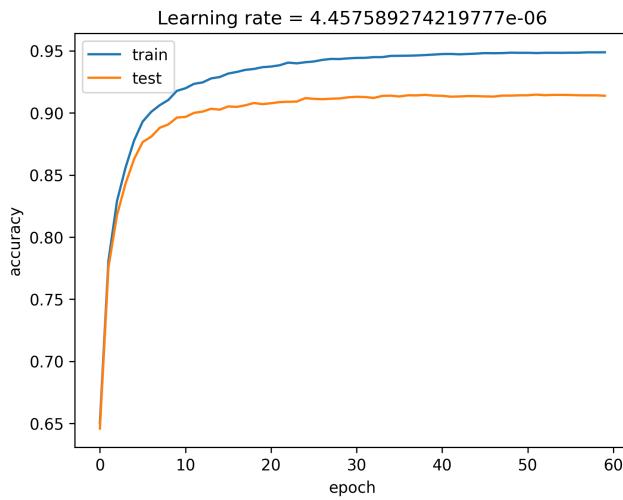
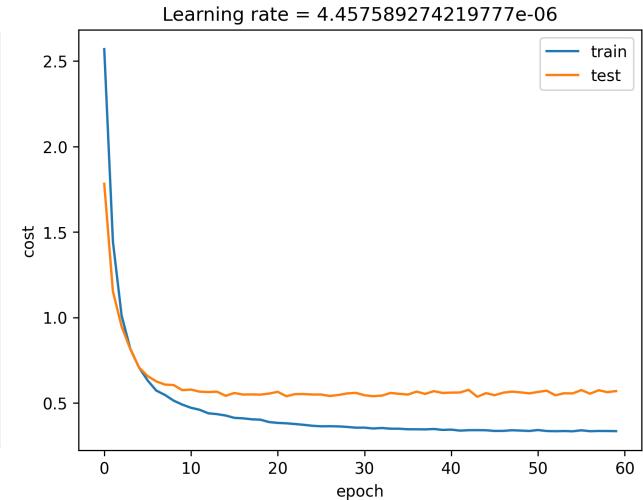


Figure 13: Confusion matrix for the best performing model



Accuracy curve



Cost curve



Some of the misclassified images