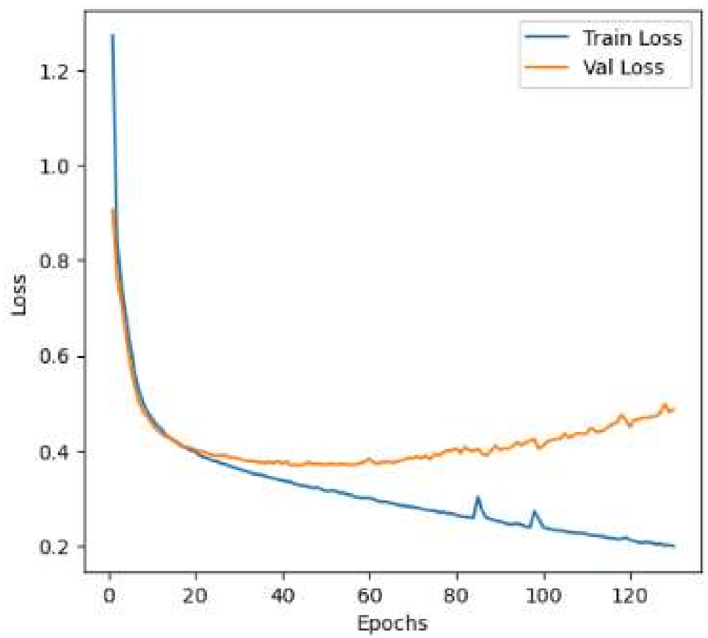
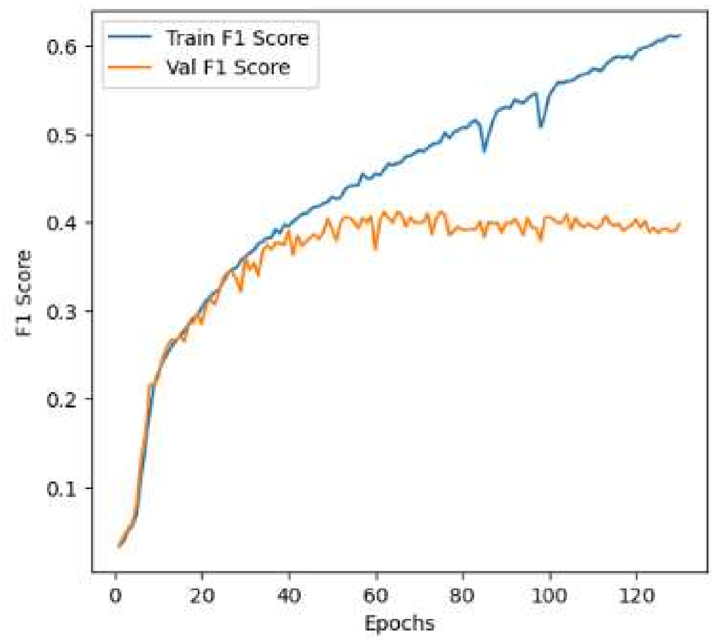
Graphs

T1 MODEL-1 FASTTEXT

Train loss, Validation loss

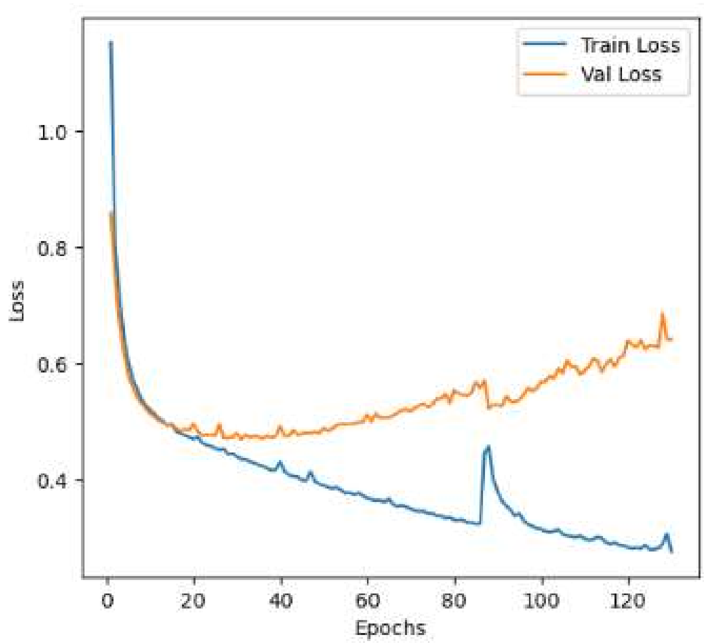


Train F1, Validation F1

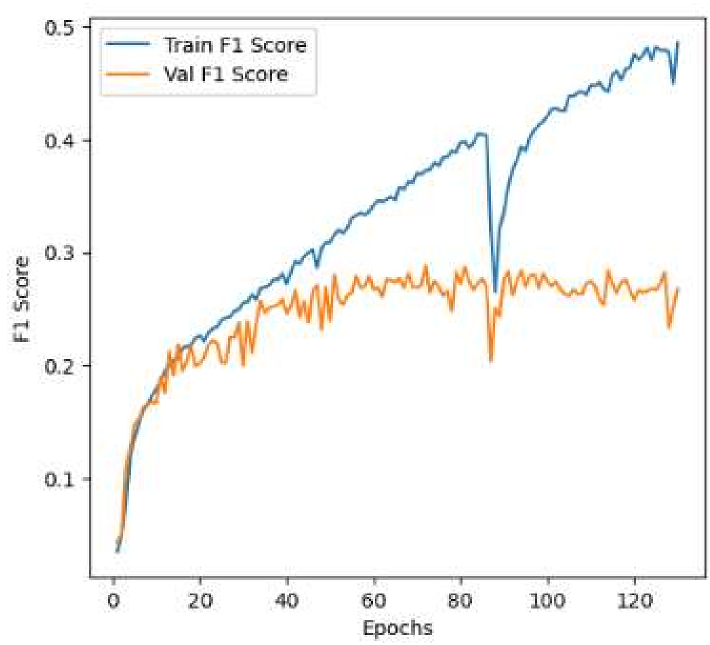


T1 MODEL-1 GLOVE

Train loss, Validation loss

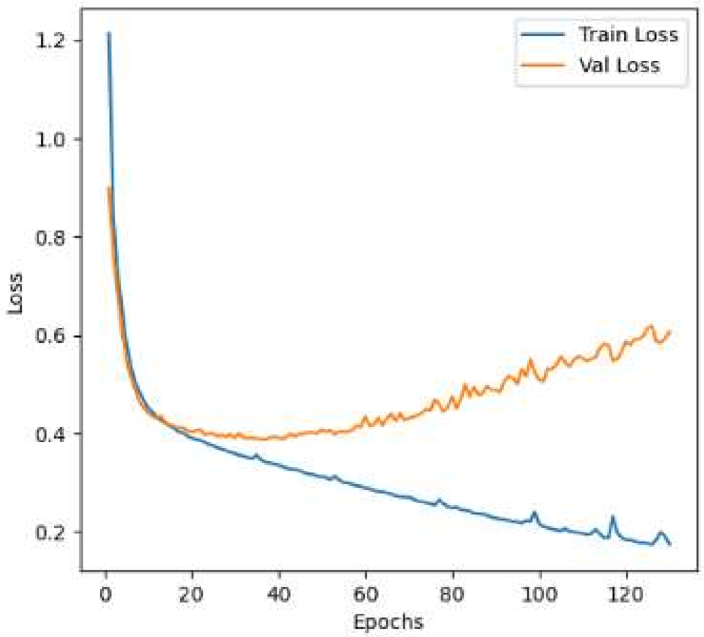


Train F1, Validation F1

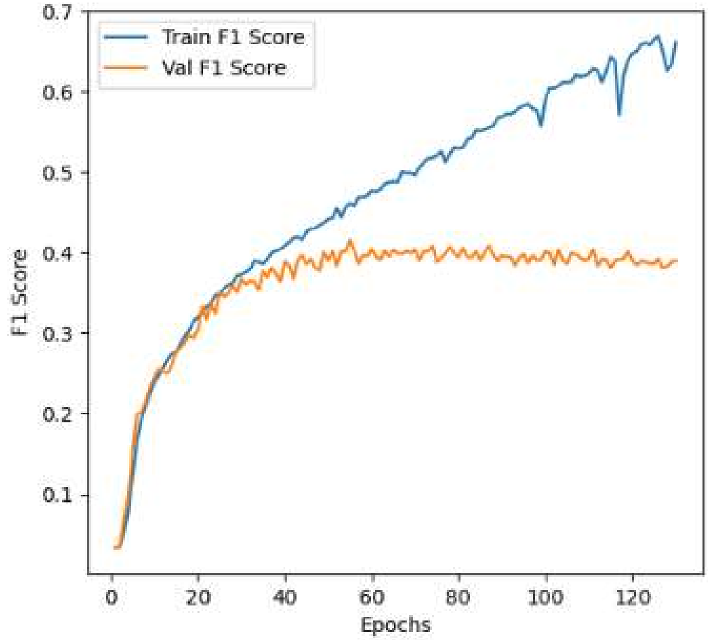


T1 MODEL-1 WORD2VEC

Train loss, Validation loss

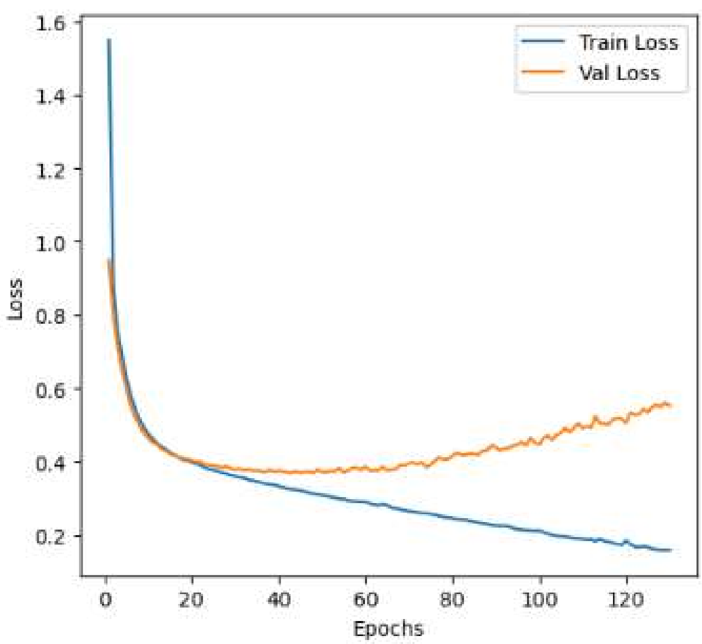


Train F1, Validation F1

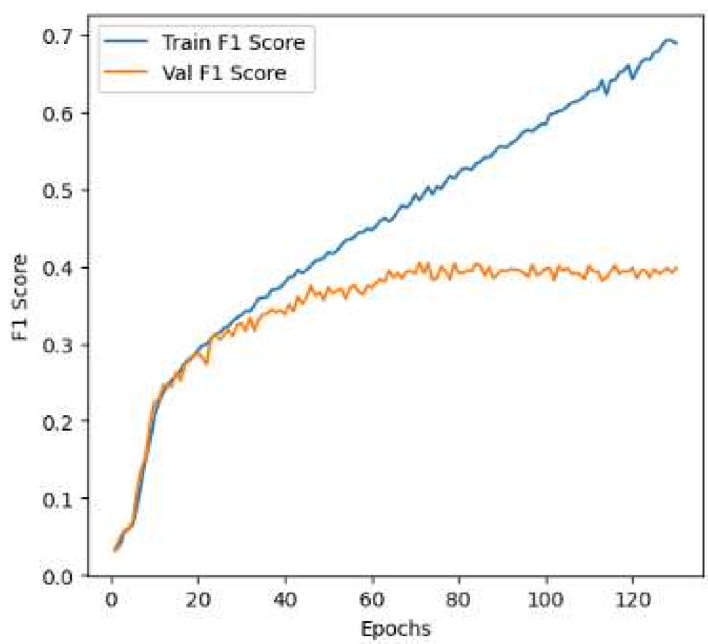


T1 MODEL-2 FASTTEXT

Train loss, Validation loss

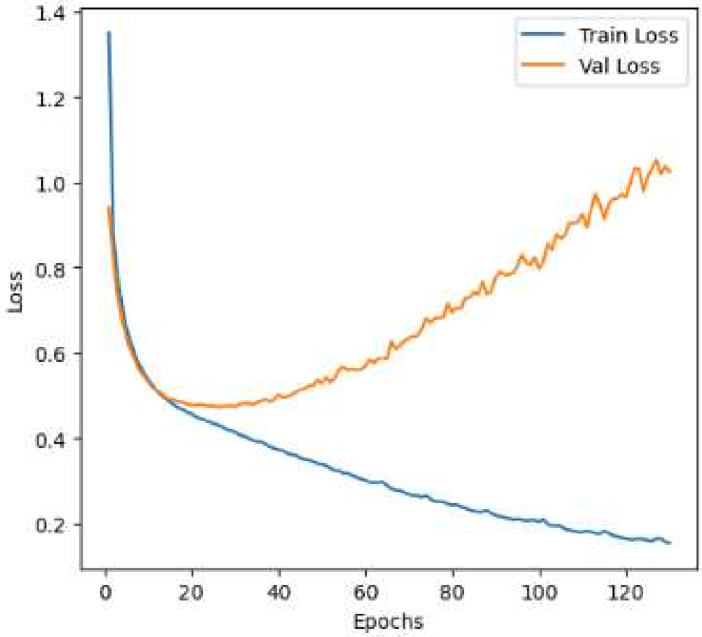


Train F1, Validation F1

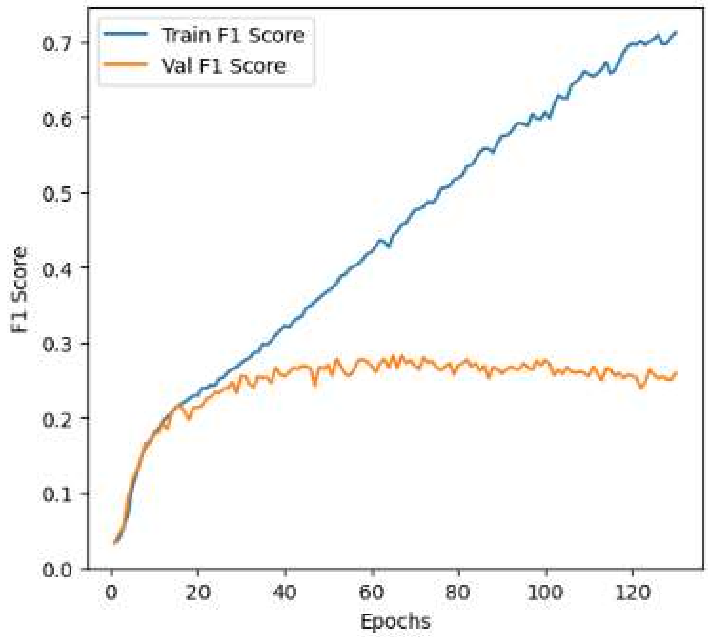


T1 MODEL-2 GLOVE

Train loss, Validation loss

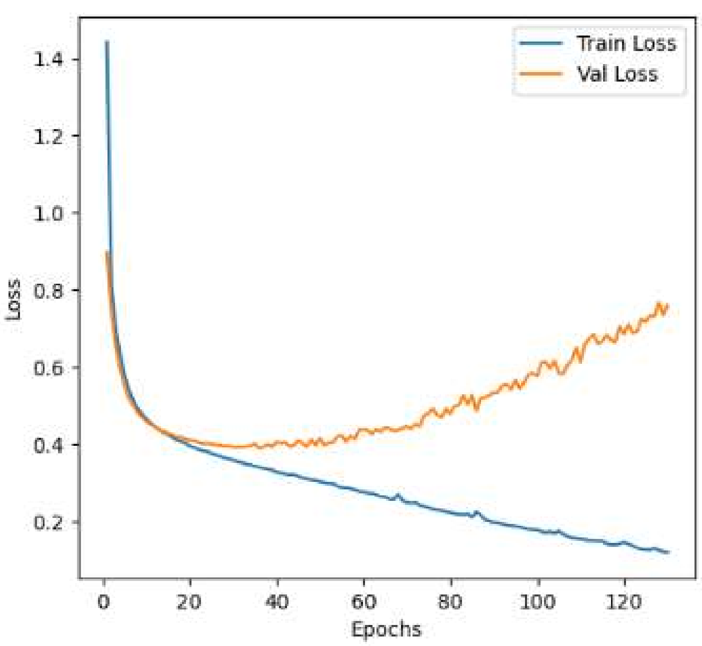


Train F1, Validation F1

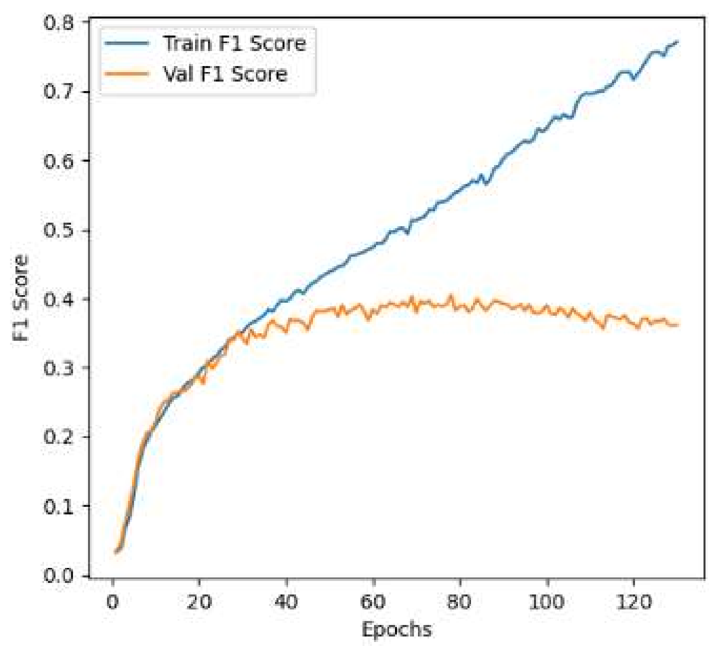


T1 MODEL-2 WORD2VEC

Train loss, Validation loss

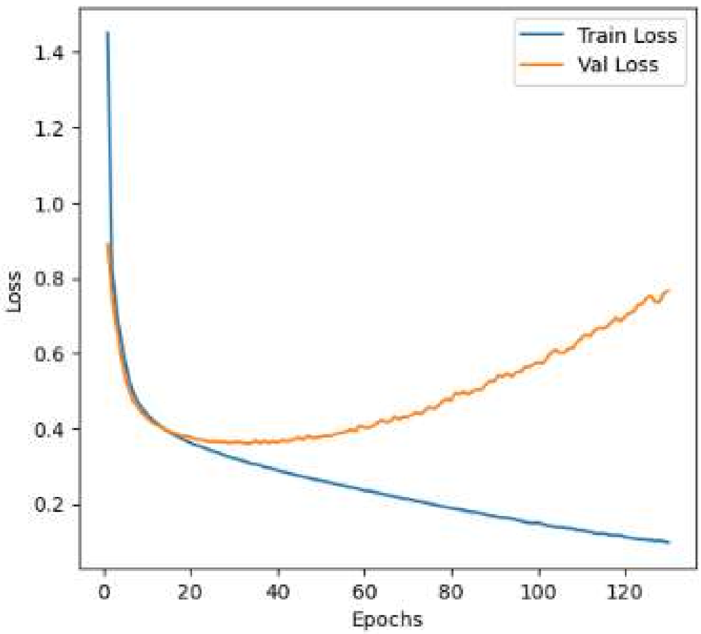


Train F1, Validation F1

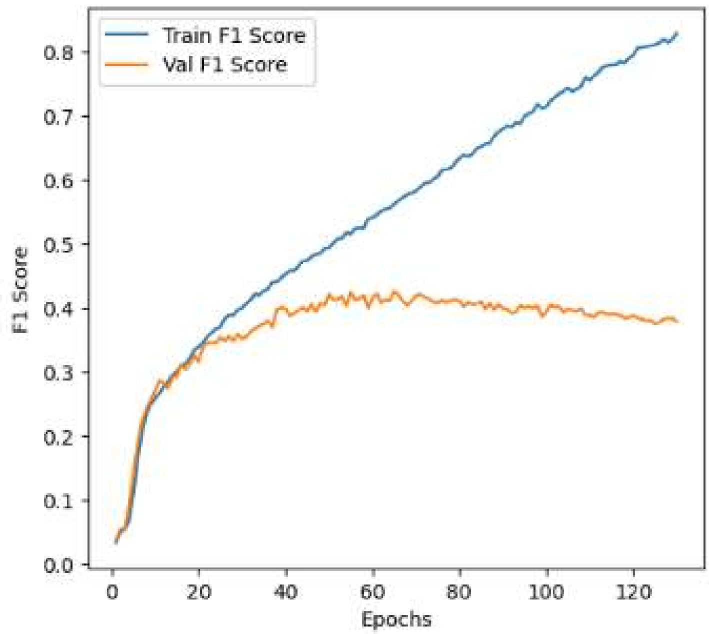


T1 MODEL-3 FASTTEXT

Train loss, Validation loss

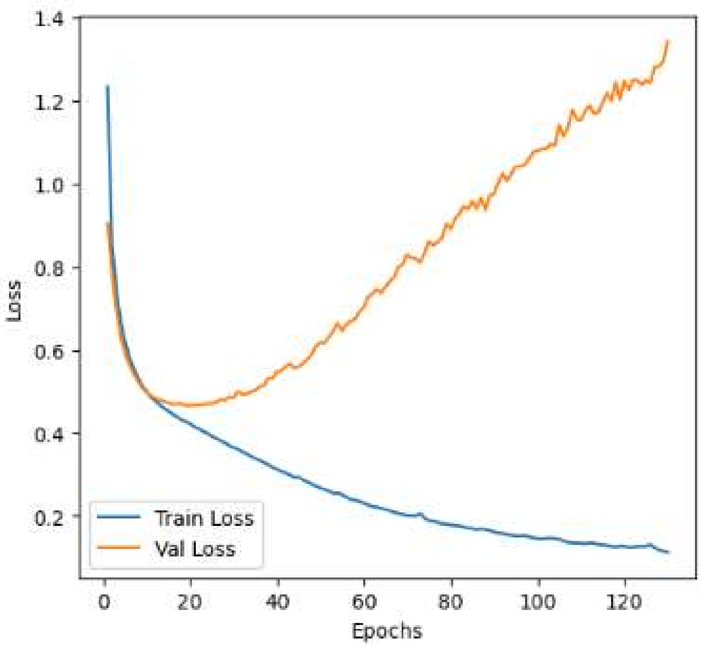


Train F1, Validation F1

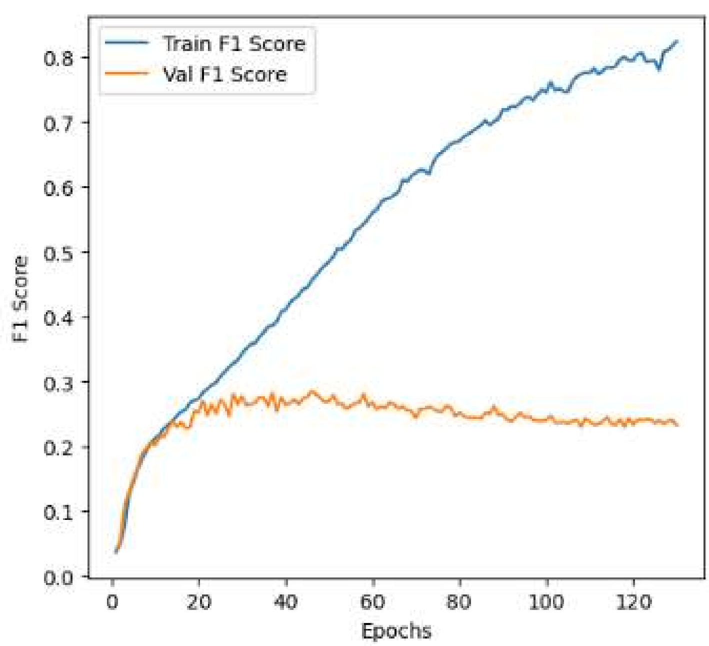


T1 MODEL-3 GLOVE

Train loss, Validation loss

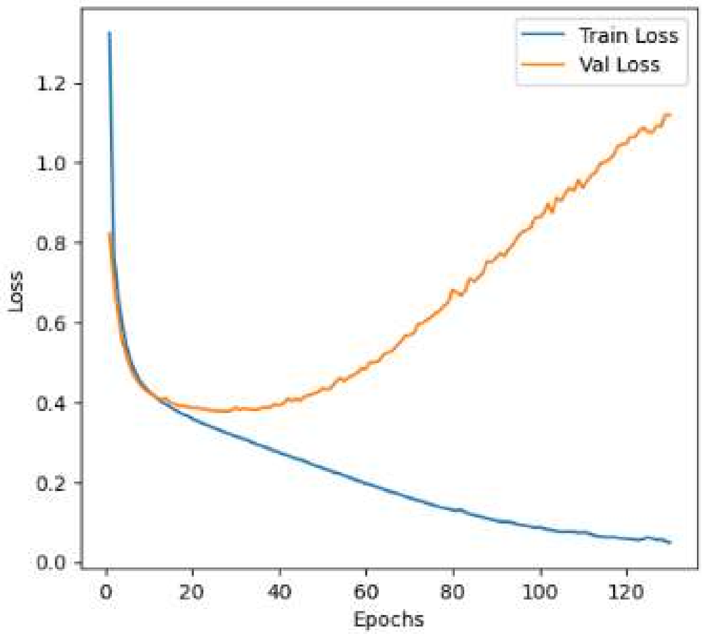


Train F1, Validation F1

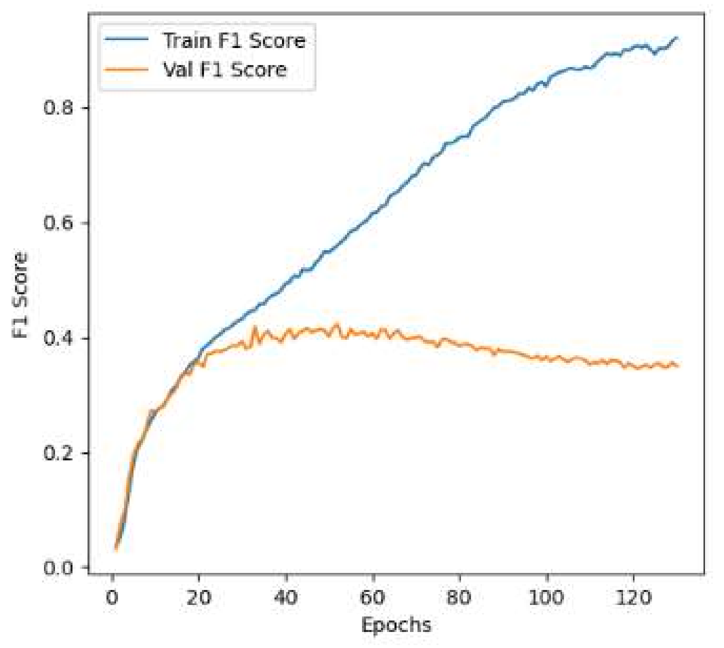


T1 MODEL-3 WORD2VEC

Train loss, Validation loss



Train F1, Validation F1

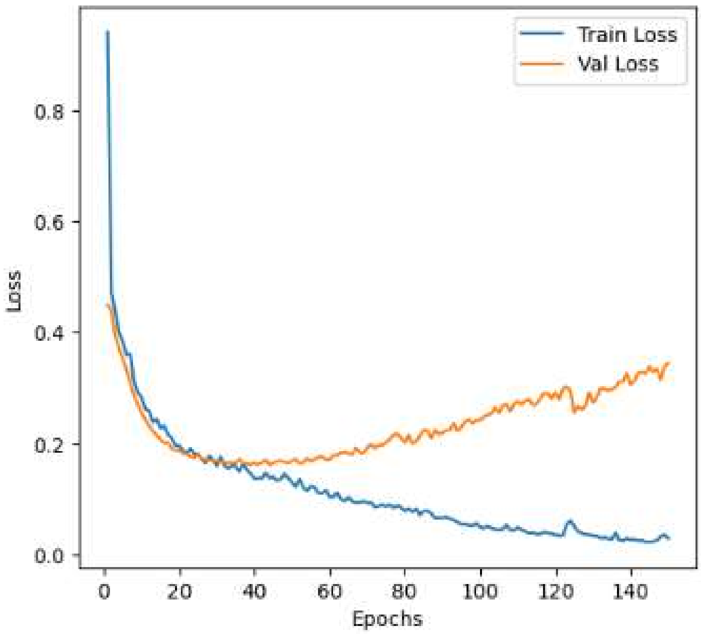


In task T1, utilizing dataset 1, the observations are as follows:

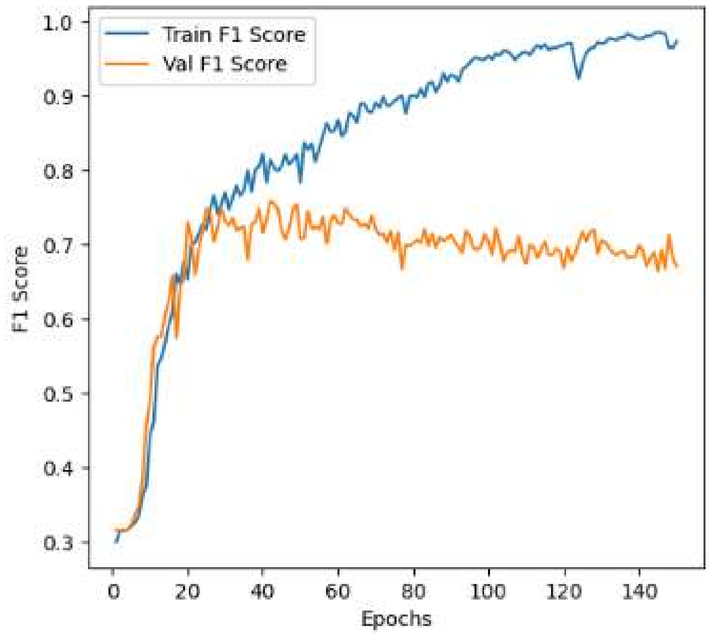
- The overall performance of RNN resulted in a lower F1 score compared to both LSTM and GRU models.
- Notably, while LSTM and GRU exhibited comparable F1 scores, RNN showed poor performance due to the simplicity of the RNN model, which may limit its capacity to learn a large number of classes effectively.
- Moreover, when considering the embedding techniques employed, word2vec emerges as the better performer than Glove and fastText.
- The Glove embedding models experienced overfitting, indicated by low training loss contrasted with notably higher validation loss. This implies that although Glove embeddings perform well during training, they struggle to generalize to new data, likely because they fail to capture the intricate semantic relationships within the dataset.

T2 MODEL-1 FASTTEXT

Train loss, Validation loss

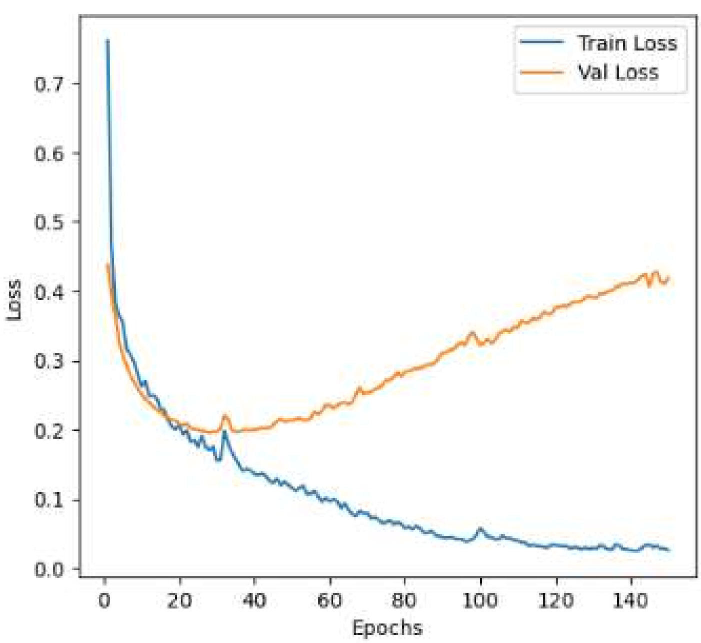


Train F1, Validation F1

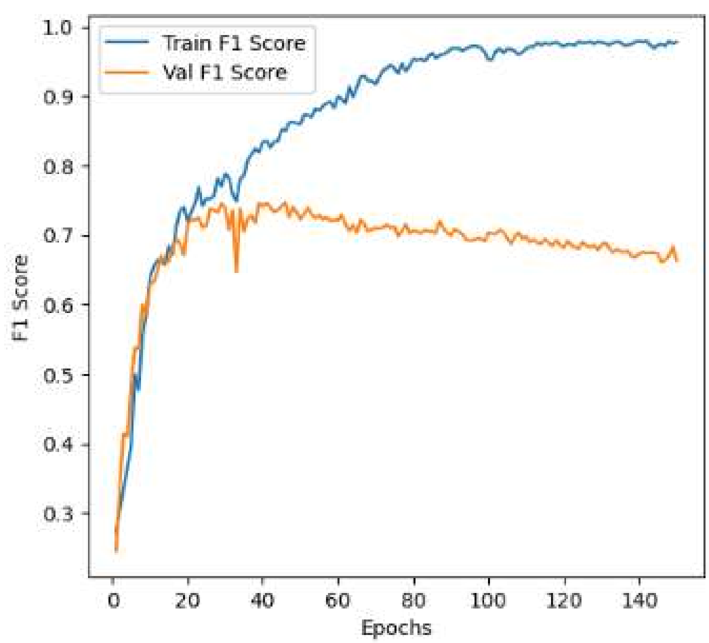


T2 MODEL-1 GLOVE

Train loss, Validation loss

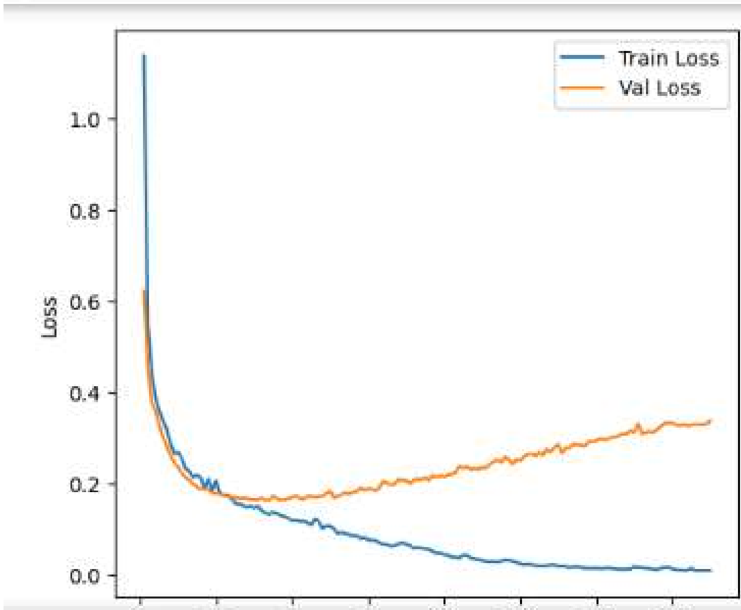


Train F1, Validation F1

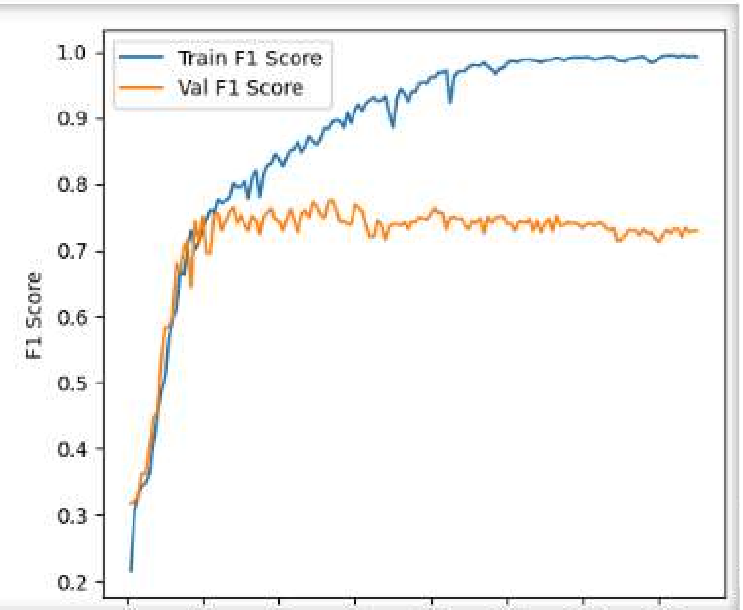


T2 MODEL-1 WORD2VEC

Train loss, Validation loss

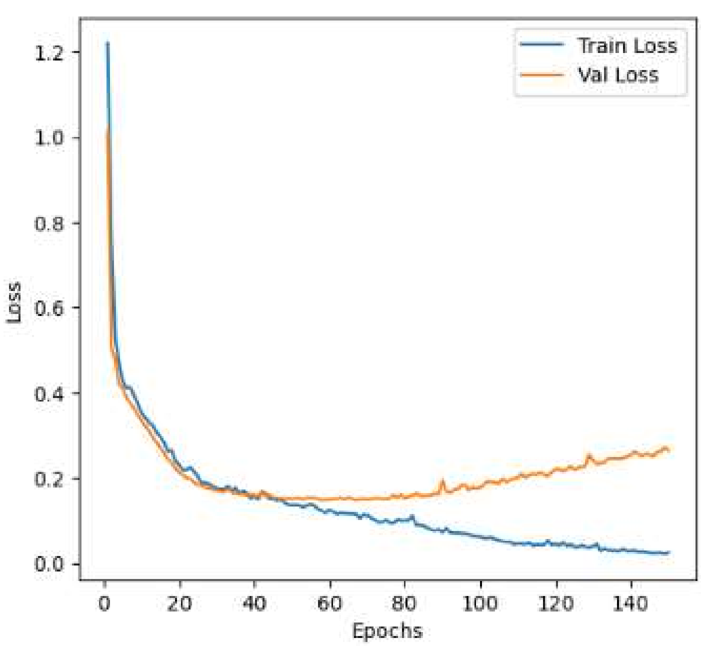


Train F1, Validation F1

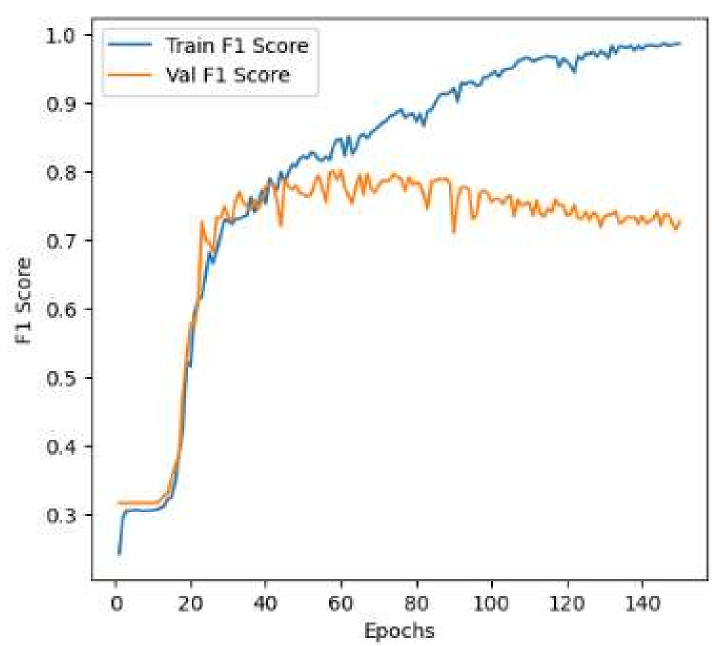


<u>T2 MODEL-2 FASTEXT</u>

Train loss, Validation loss

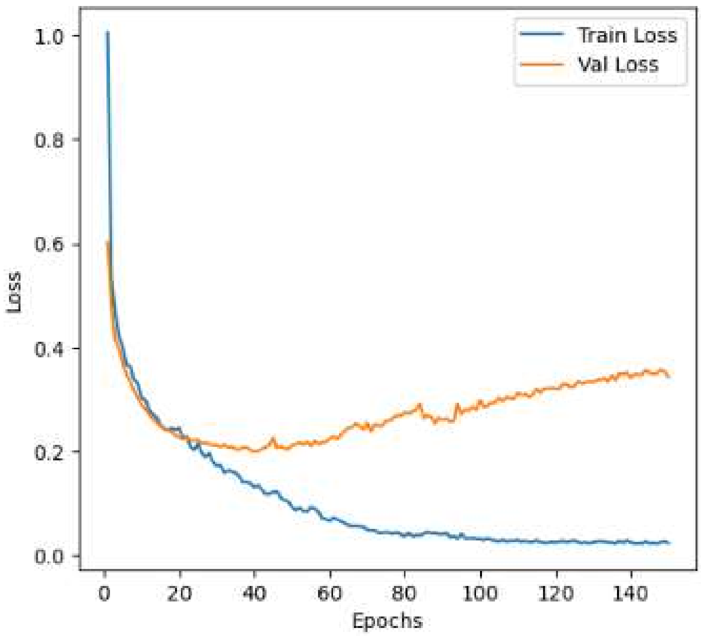


Train F1, Validation F1

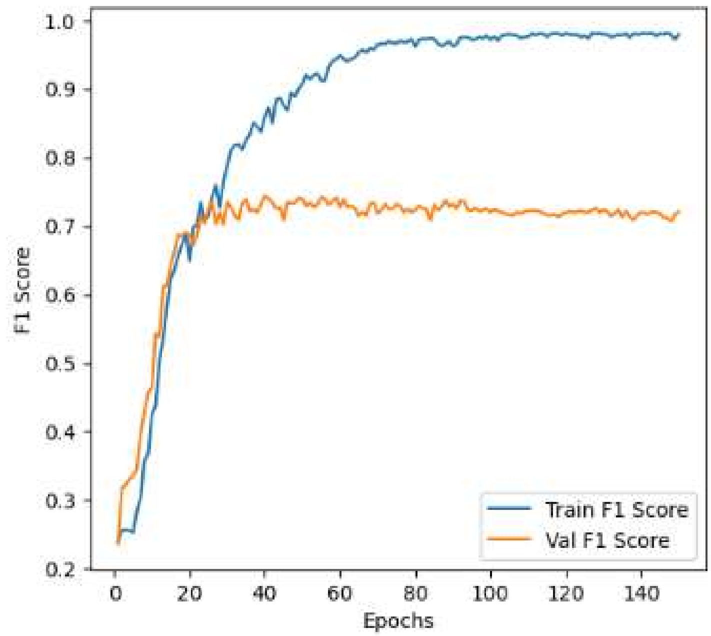


<u>T2 MODEL-2 GLOVE</u>

Train loss, Validation loss

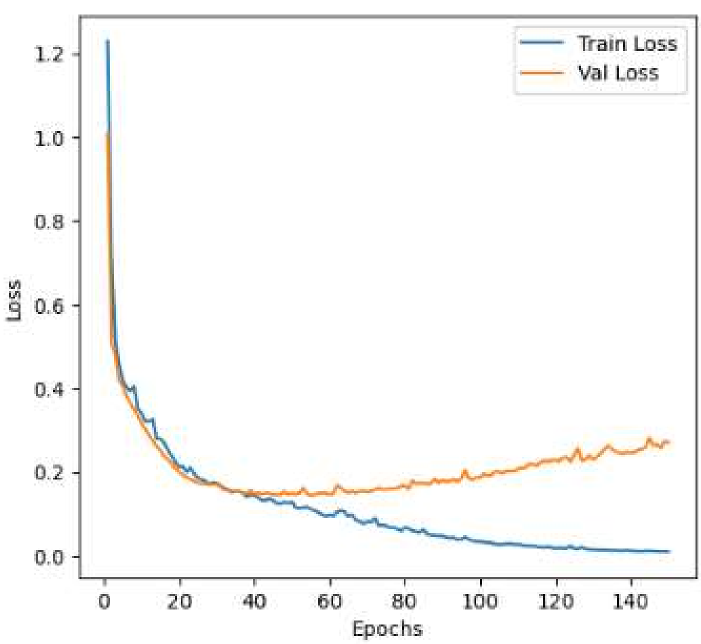


Train F1, Validation F1

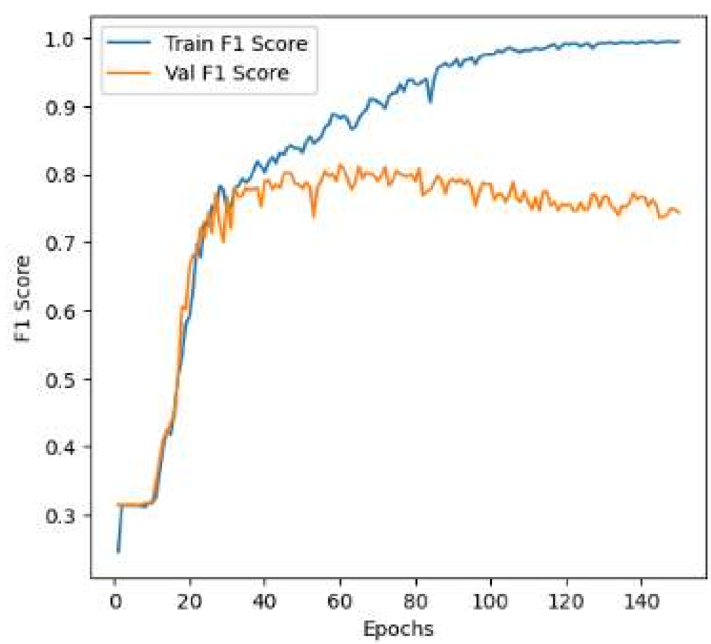


T2 MODEL-2 WORD2VEC

Train loss, Validation loss

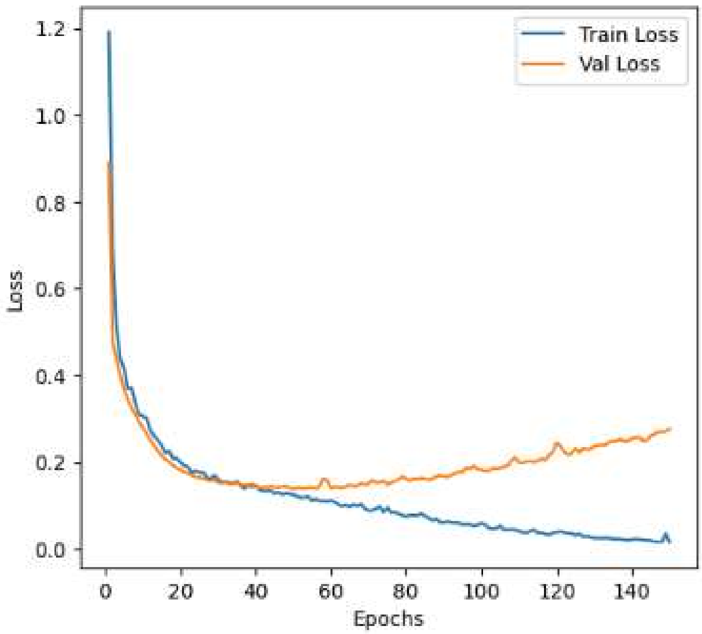


Train F1, Validation F1

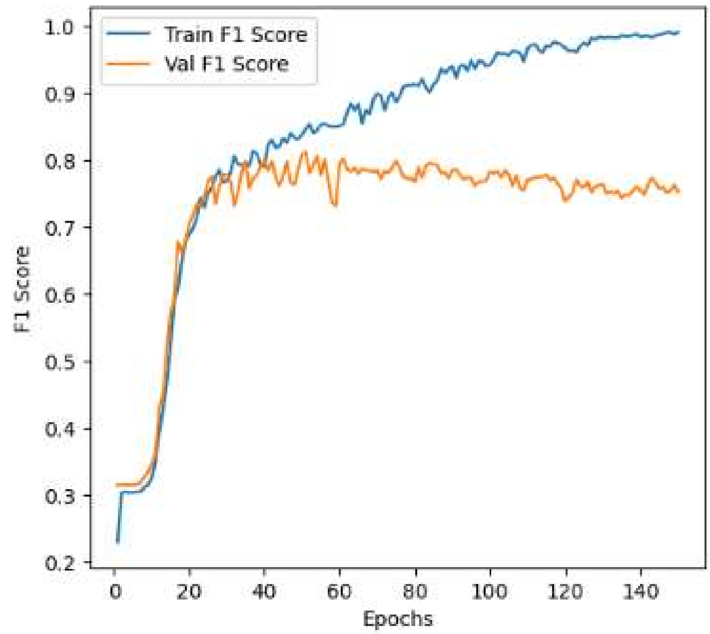


T2 MODEL-3 FASTTEXT

Train loss, Validation loss

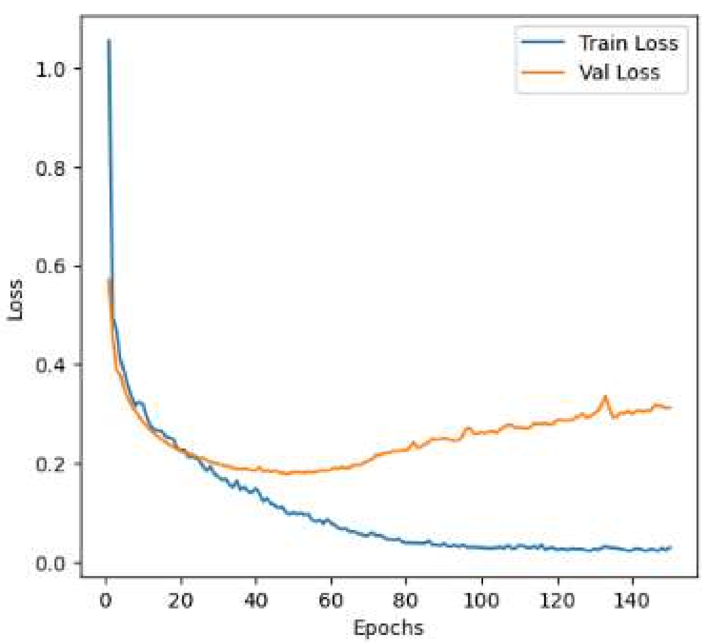


Train F1, Validation F1

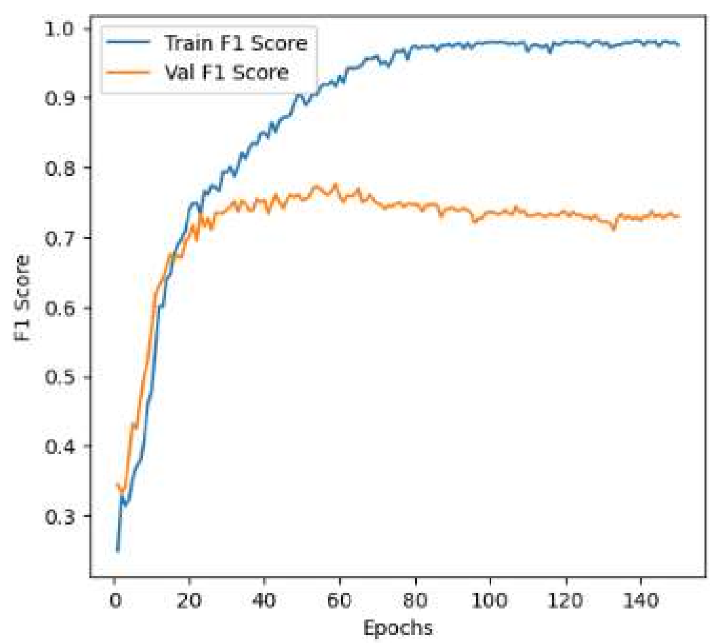


T2 MODEL-3 GLOVE

Train loss, Validation loss

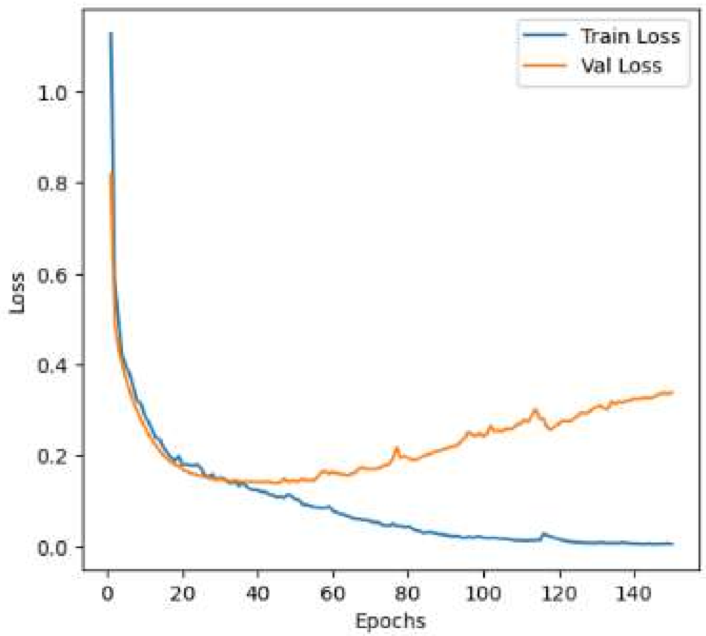


Train F1, Validation F1

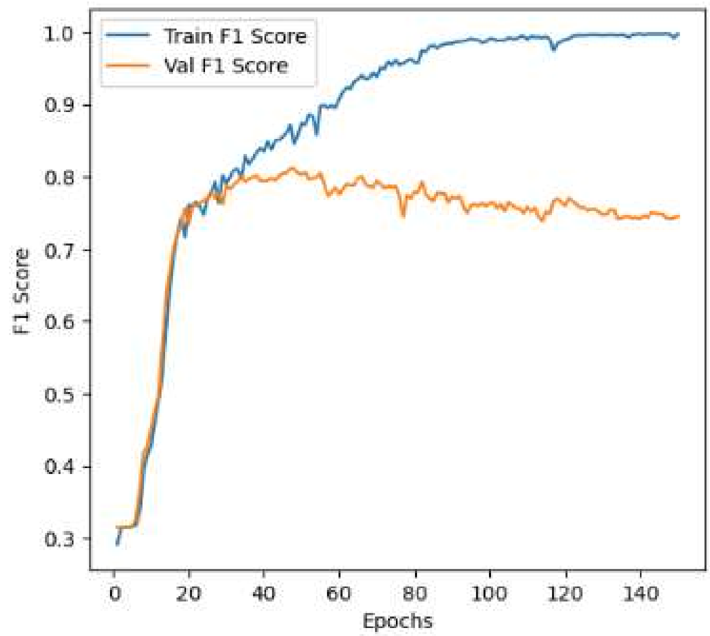


T2 MODEL–3 WORD2VEC

Train loss, Validation loss

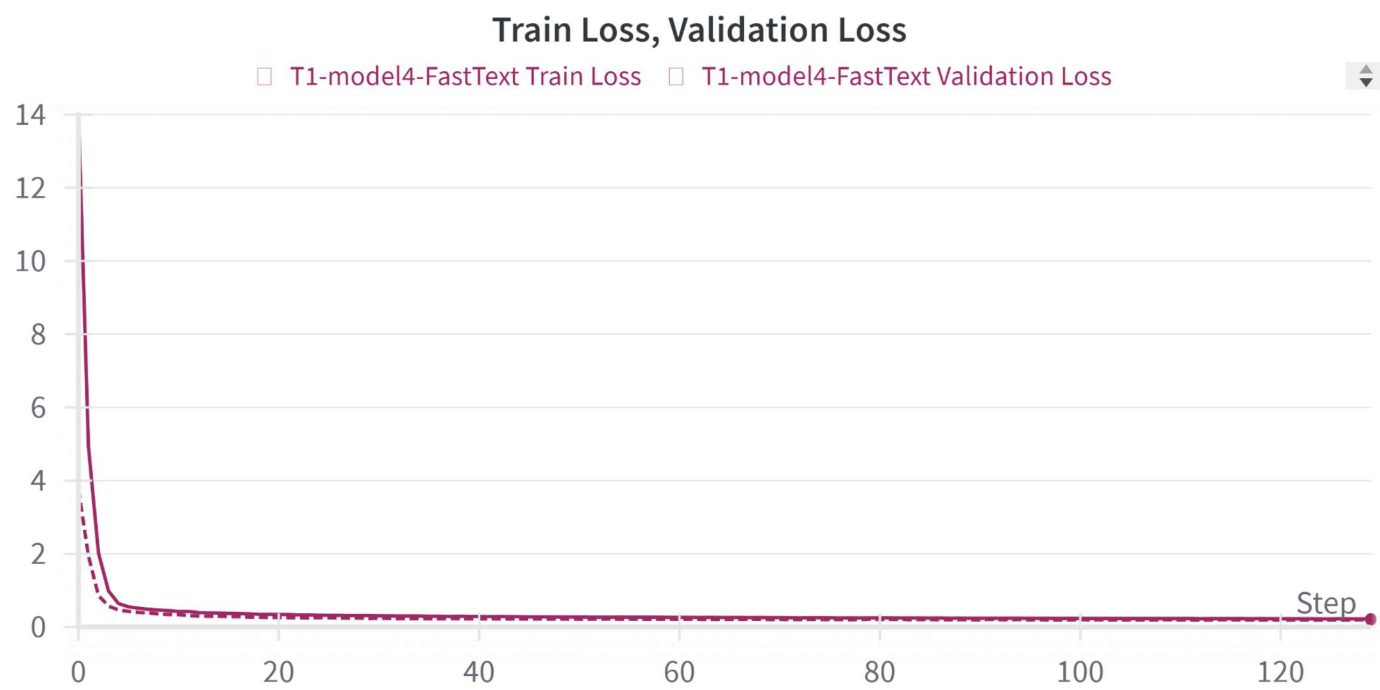


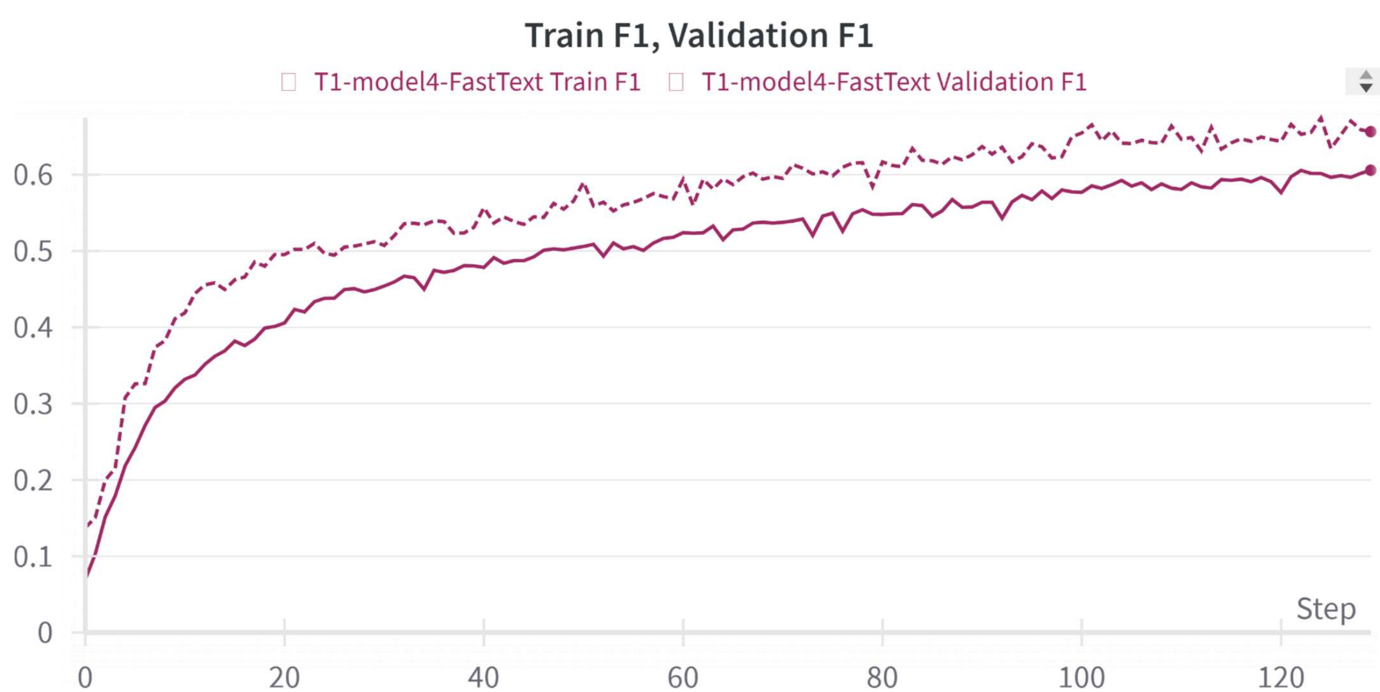
Train F1, Validation F1

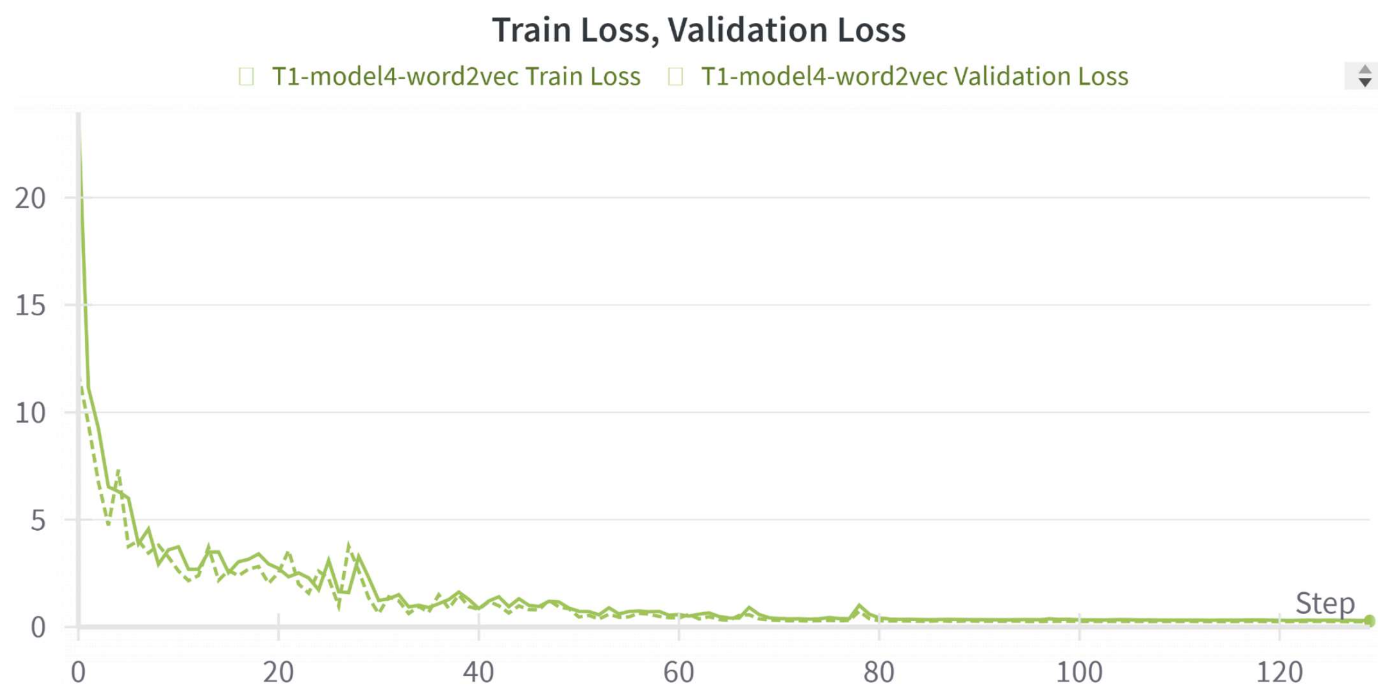


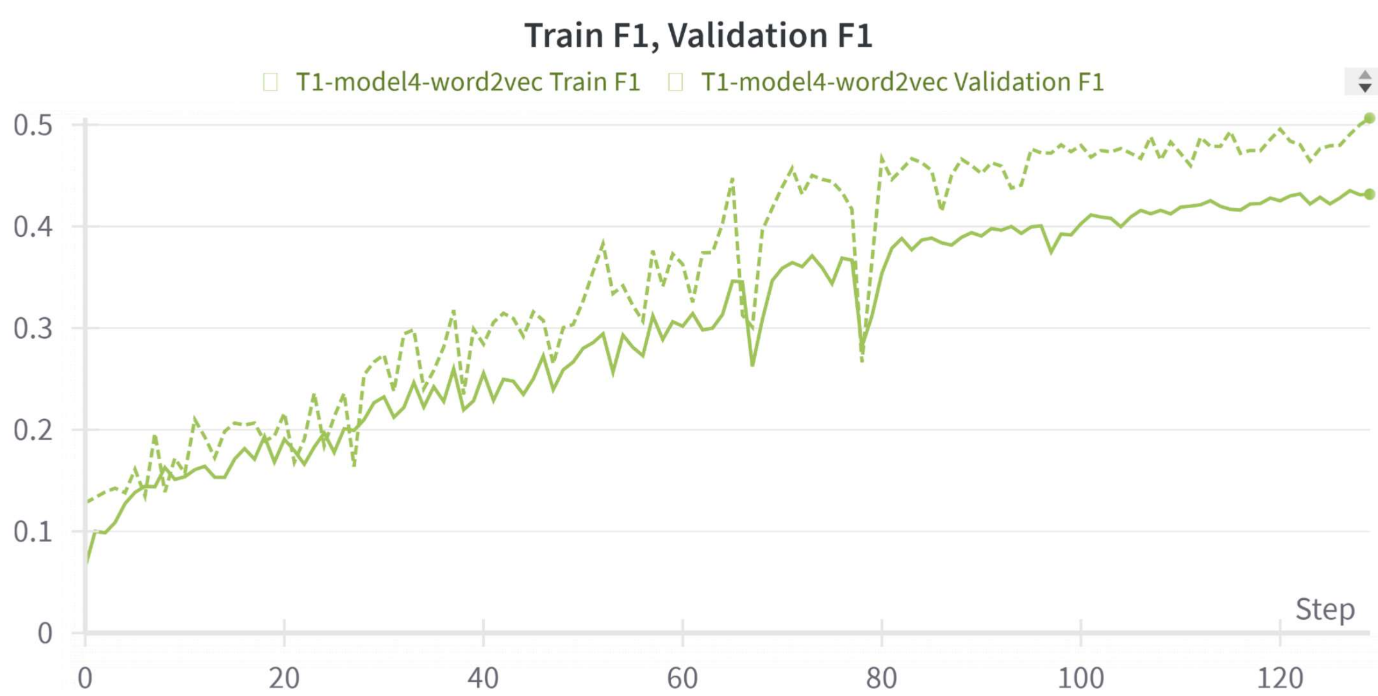
In Task 2, employing dataset 2, it's evident that the GloVe model exhibits a higher degree of overfitting compared to the others. Despite having only three classes—RNN, LSTM, and GRU—their performances were remarkably similar, showcasing comparable effectiveness in handling the task's complexities.

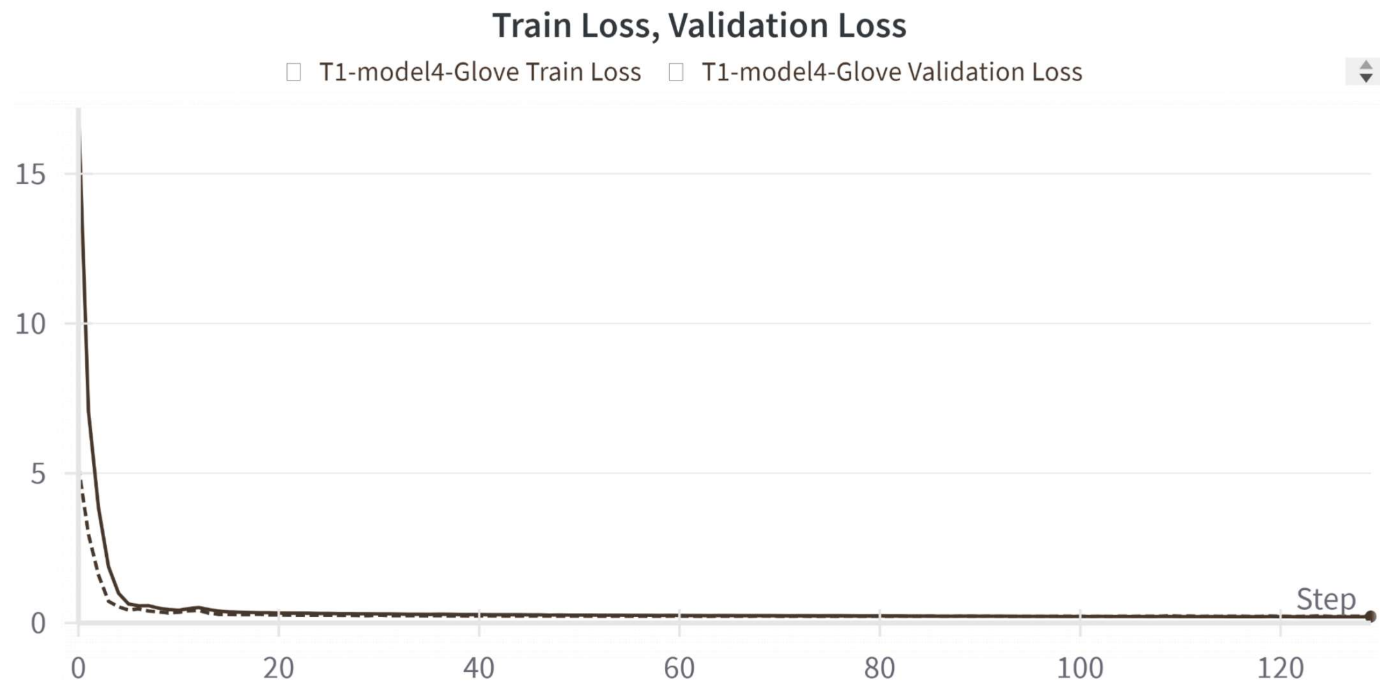
T1_MODEL-4

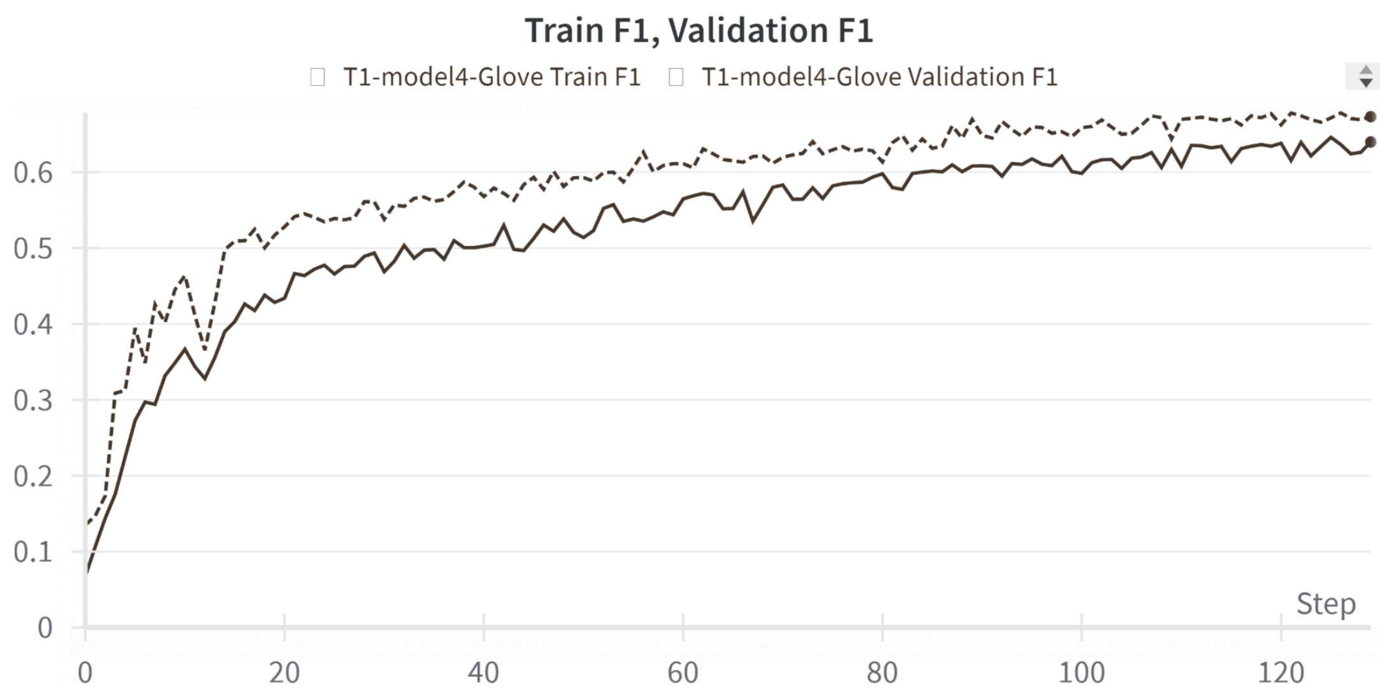

Train Loss, Validation Loss
T1-model4-FastText Train Loss
T1-model4-FastText Validation Loss
Step


Train F1, Validation F1
T1-model4-FastText Train F1
T1-model4-FastText Validation F1
Step


Train Loss, Validation Loss
T1-model4-word2vec Train Loss
T1-model4-word2vec Validation Loss
20
15
10
5
0
Step
0
20
40
60
80
100
120

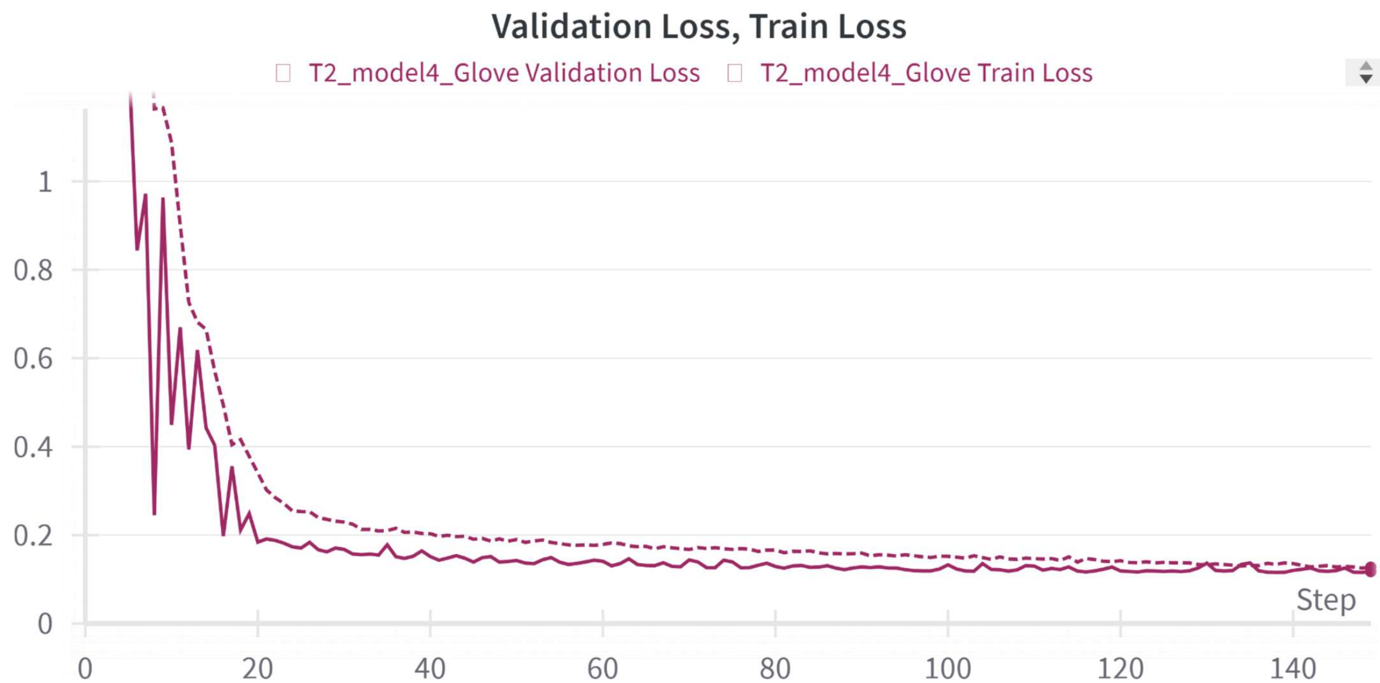

Train F1, Validation F1
T1-model4-word2vec Train F1
T1-model4-word2vec Validation F1
0.5
0.4
0.3
0.2
0.1
0
Step
0
20
40
60
80
100
120

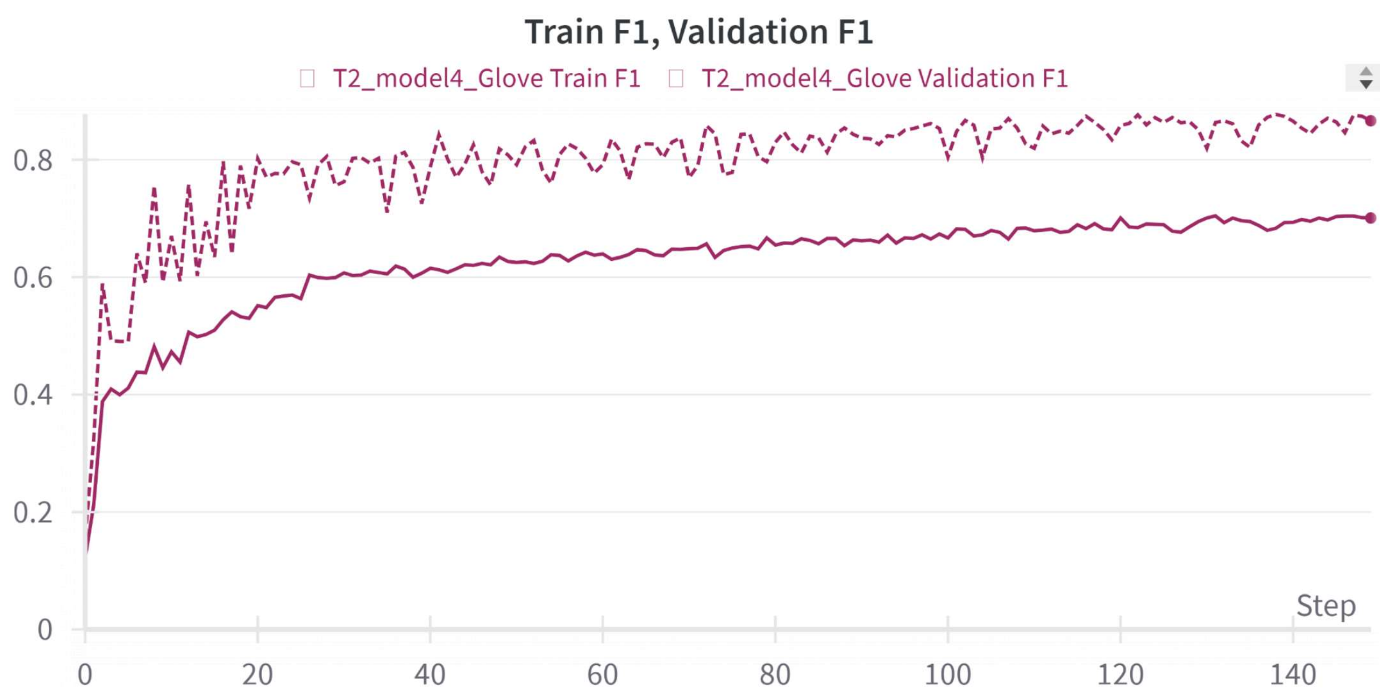


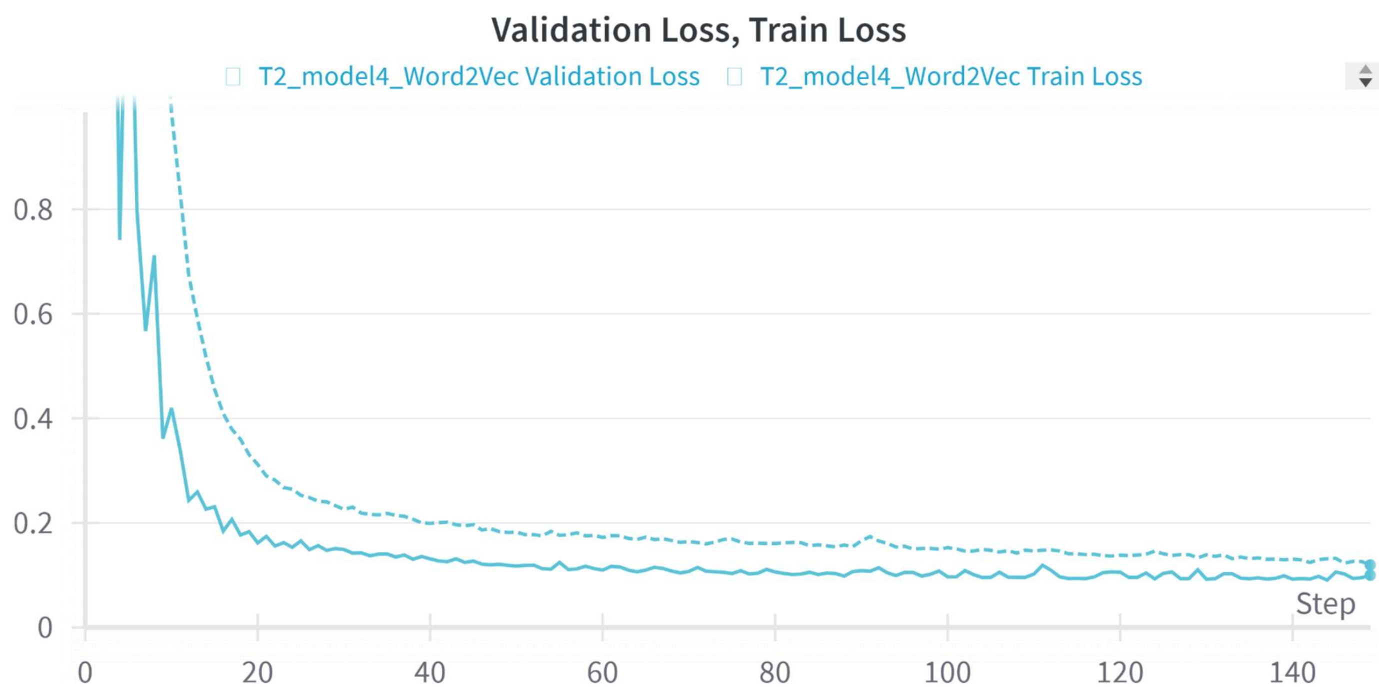


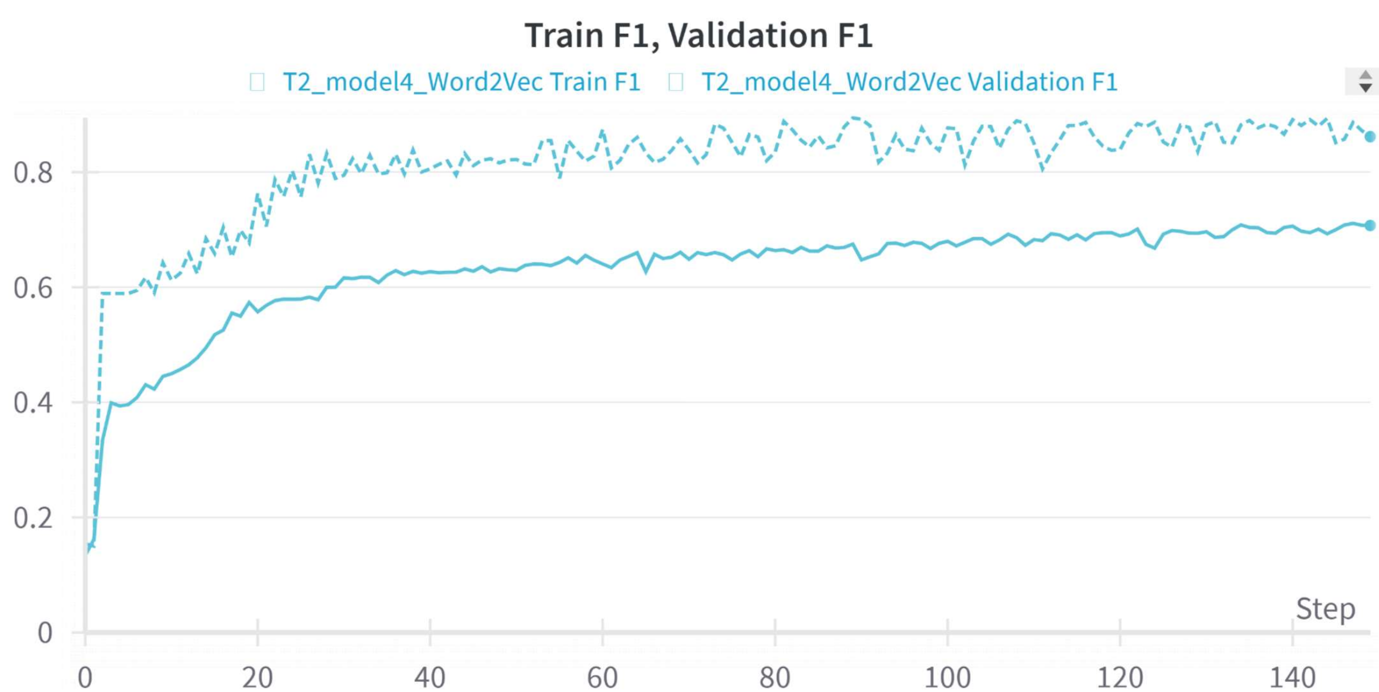
In the context of task t2, the Word2Vec embedding model demonstrates a challenge as it yields notably high training and validation losses, suggesting the presence of substantial biases and a limited capacity to effectively understand the underlying data patterns. In contrast, both the FastText and GloVe embeddings showcase more promising results. These two models exhibit similar trends in the convergence of loss, F1 score, and accuracy during training. Notably, GloVe stands out by achieving superior performance on the validation set, emphasising its effectiveness in capturing and generalising from the task-specific data.

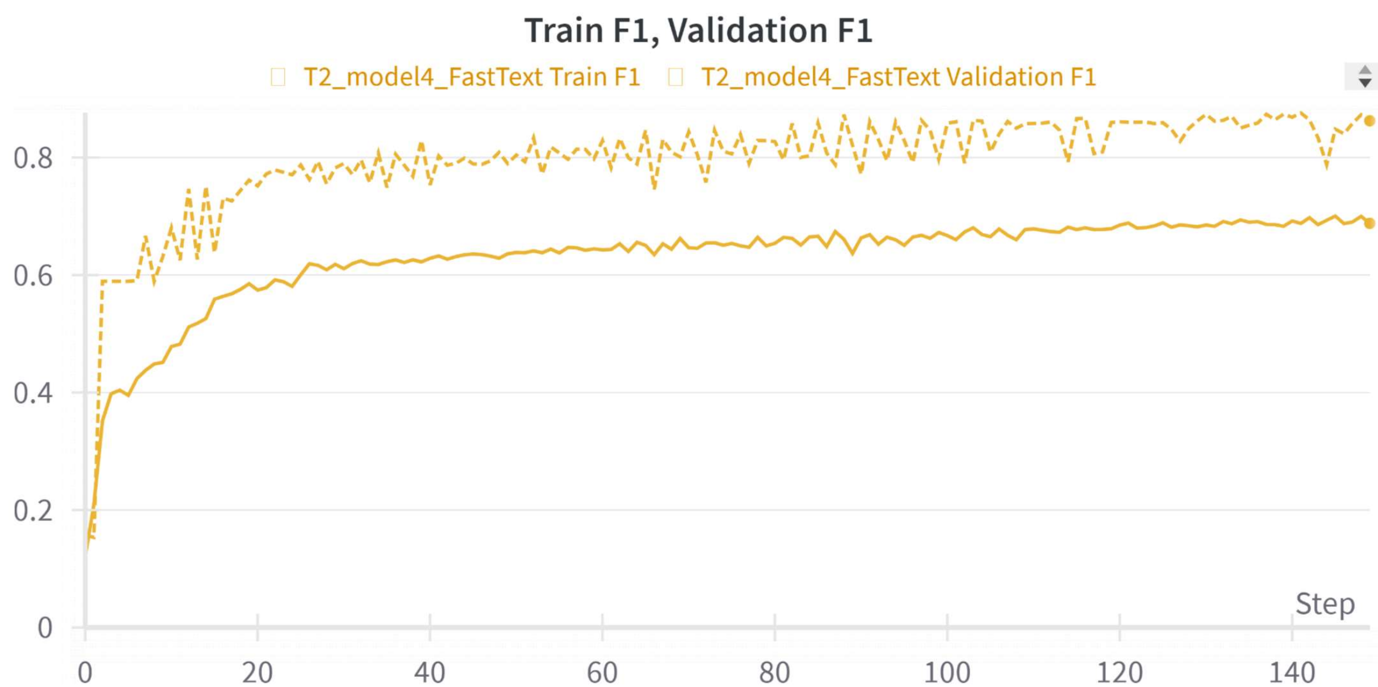
T2_MODEL4






Validation Loss, Train Loss
T2_model4_Word2Vec Validation Loss
T2_model4_Word2Vec Train Loss
0.8
0.6
0.4
0.2
0
Step
0
20
40
60
80
100
120
140


Train F1, Validation F1
T2_model4_Word2Vec Train F1
T2_model4_Word2Vec Validation F1
0.8
0.6
0.4
0.2
0
Step
0
20
40
60
80
100
120
140



NOTE - Training score is in solid line, and validation score is in dash line

All tested embeddings demonstrate remarkably similar performance, exhibiting nearly identical convergence in terms of loss, accuracy, and F1 scores. However, it is noteworthy that FastText slightly outperforms the other embeddings. Additionally, the absence of overfitting is indicated by the minimal difference between training and validation loss throughout the training process

Task 1

Model No	Embedding_used	Accuracy	Macro_F1
1	Word2Vec	0.8618	0.3558189079041109
1	Glove	0.8491	0.23224886162702643
1	Fasttext	0.8789	0.4028700936231574
2	Word2Vec	0.8634	0.3547655557172286
2	Glove	0.8332	0.2362976720430671
2	Fasttext	0.8756	0.3948665838914022
3	Word2Vec	0.8446	0.32748062060098565
3	Glove	0.8307	0.21406890878378082
3	Fasttext	0.8627	0.3819970956544048
4	Word2Vec	0.9282	0.4691779879215485
4	Glove	0.9568	0.634601841513961
4	**Fasttext**	**0.9644**	**0.6743593325786298**

F1 Score for 'O': 0.95782214
F1 Score for 'B_COURT': 0.87761194
F1 Score for 'I_COURT': 0.85298869
F1 Score for 'B_PETITIONER': 0.61538462
F1 Score for 'I_PETITIONER': 0.77777778
F1 Score for 'B_RESPONDENT': 0.0
F1 Score for 'I_RESPONDENT': 0.0
F1 Score for 'B_JUDGE': 0.4
F1 Score for 'I_JUDGE': 0.57142857
F1 Score for 'B_DATE': 0.67148014
F1 Score for 'I_DATE': 0.68292683
F1 Score for 'B_ORG': 0.5620438
F1 Score for 'I_ORG': 0.38554217
F1 Score for 'B_GPE': 0.81313131
F1 Score for 'I_GPE': 0.84903226
F1 Score for 'B_STATUTE': 0.86767896
F1 Score for 'I_STATUTE': 0.8592233
F1 Score for 'B_PROVISION': 0.71473354
F1 Score for 'I_PROVISION': 0.87587822
F1 Score for 'B_PRECEDENT': 0.67
F1 Score for 'I_PRECEDENT': 0.79884226
F1 Score for 'B_CASE_NUMBER': 0.41025641
F1 Score for 'I_CASE_NUMBER': 0.43037975

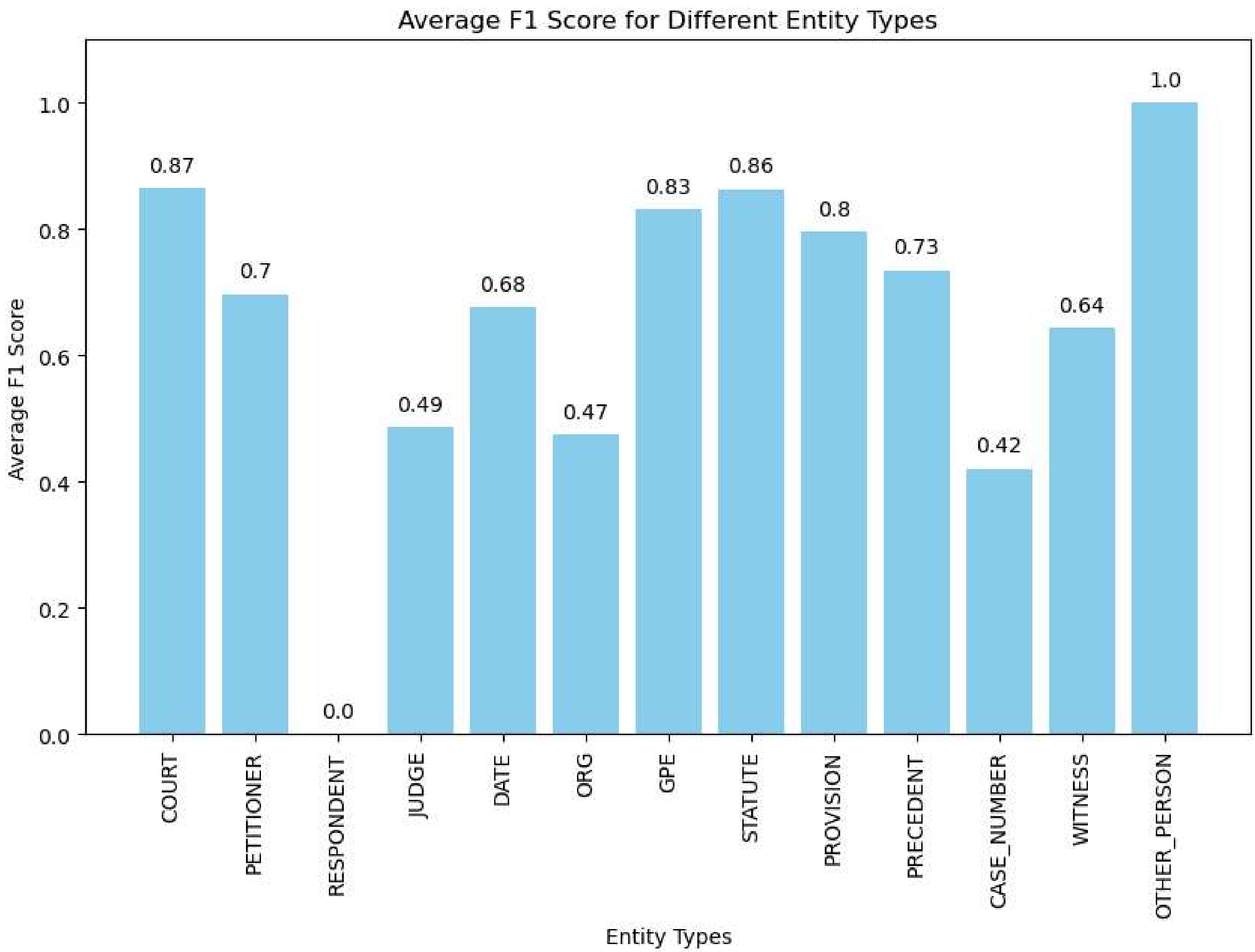
F1 Score for 'B_WITNESS': 0.62937063
F1 Score for 'I_WITNESS': 0.65625
F1 Score for 'B_OTHER_PERSON': 1.0
F1 Score for 'I_OTHER_PERSON': 1.0

AVERAGE-13 class

[

1. **Average F1 Score for 'COURT': 0.865300315-**
2. **Average F1 Score for 'PETITIONER': 0.6965812-**
3. **Average F1 Score for 'RESPONDENT': 0.0 -**
4. **Average F1 Score for 'JUDGE': 0.485714285 -**
5. **Average F1 Score for 'DATE': 0.6772034849999999-**
6. **Average F1 Score for 'ORG': 0.473792985-**
7. **Average F1 Score for 'GPE': 0.8310817850000001-**
8. **Average F1 Score for 'STATUTE': 0.8634511300000001-**
9. **Average F1 Score for 'PROVISION': 0.79530588-**
10. **Average F1 Score for 'PRECEDENT': 0.7344211300000001**
11. **Average F1 Score for 'CASE_NUMBER': 0.42031808000000004-**
12. **Average F1 Score for 'WITNESS': 0.642810315-**
13. **Average F1 Score for 'OTHER_PERSON': 1.0-**

]



Task 2

Model No	Embedding_used	Accuracy	Macro_F1
1	Word2Vec	0.9062	0.6999402532919099
1	Glove	0.8849	0.6017197129354663
1	Fasttext	0.9098	0.688587030489776
2	Word2Vec	0.9149	0.709729054763831
2	Glove	0.9021	0.6708463379432797
2	Fasttext	0.9143	0.7154978149353443
3	Word2Vec	0.9135	0.7199227729549523

3	Glove	0.9038	0.6610621281734312
3	Fasttext	0.9191	0.7335586703984592
4	**Word2Vec**	**0.9849**	**0.836901469083168**
4	Glove	0.9796	0.8319482704772927
4	Fasttext	0.9786	0.8216757024437102

F1 Best Model Classwise:
F1 Score for 'O': 0.958638
F1 Score for 'B': 0.67002519
F1 Score for 'I': 0.55584416
F1 Score for '<START>': 1.0
F1 Score for '<STOP>': 1.0

Snippets of Dataset

T1—

During the dataloader creation process, an additional layer of preprocessing is integrated to enhance the quality of the dataset. Specifically, this involves the elimination of stopwords along with their corresponding labels, particularly when the label does not pertain to the 'O' class. Furthermore, a crucial step involves the normalization of text to lowercase. These preprocessing techniques collectively contribute to refining the dataset, ultimately aiming to improve the overall suggestions generated by the model. Because we have 27 classes, which implies having higher model complexity and parameter so normlization will reduce complexity.


```
"0": {
  "text": "Clause 18(1), (2) and (3)\n(a) & (b) were transposed in Article 23 of the Draft Constitution of India.",
  "labels": [
    "O",
    "O",
    "O",
    "O",
    "O",
    "O",
    "O",
    "O",
    "O",
    "O",
    "B_PROVISION",
    "I_PROVISION",
    "O",
    "O",
    "B_STATUTE",
    "I_STATUTE",
    "I_STATUTE",
    "I_STATUTE",
    "O"
  ]
```


"0": {
"text": "Clause 18(1), (2) and (3)\n(a) & (b) were transposed in Article 23 of the Draft Constitution of India.",
"labels": [
"O",

"O",
"O",
"O",
"O",
"O",
"O",
"O",
"O",
"O",
"B_PROVISION",
"I_PROVISION",
"O",
"O",
"B_STATUTE",
"I_STATUTE",
"I_STATUTE",
"I_STATUTE",
"O"
]

"6": {
"text": "4. Under Schedule I of the Calcutta Municipal Act, 1923, as also in Schedule I of the Calcutta Municipal Act, 1951 (hereinafter referred to as the 'Act') as originally enacted, \"Calcutta\" has been defined.",
"labels": [
"O",
"B_PROVISION",
"I_PROVISION",
"O",
"O",
"B_STATUTE",
"I_STATUTE",
"I_STATUTE",
"I_STATUTE",
"O",
"O",
"O",
"B_PROVISION",
"I_PROVISION",
"O",
"O",
"B_STATUTE",
"I_STATUTE",
"I_STATUTE",
"I_STATUTE",
"O",
"O",
"O",
"O",
"O",
"O",
"O",
"O",
"O",
"B_GPE",
"O",
"O",
"O",
"O"
]

"6": {
"text": "4. Under Schedule I of the Calcutta Municipal Act, 1923, as also in Schedule I of the Calcutta Municipal Act, 1951 (hereinafter referred to as the 'Act') as originally enacted, \"Calcutta\" has been defined.",
"labels": [
"O",

"B_PROVISION",
"I_PROVISION",
"O",
"O",
"B_STATUTE",
"I_STATUTE",
"I_STATUTE",
"I_STATUTE",
"O",
"O",
"O",
"B_PROVISION",
"I_PROVISION",
"O",
"O",
"B_STATUTE",
"I_STATUTE",
"I_STATUTE",
"I_STATUTE",
"O",
"O",
"O",
"O",
"O",
"O",
"O",
"O",
"O",
"B_GPE",
"O",
"O",
"O",
"O"
]

T2—

No additional preprocessing was required in this case because we have 3 classes hence lower complexity so higher information will help to learning complex features.


```
"2": {
  "text": "Easy to start up and does not overheat as much as other laptops .",
  "labels": [
    "O",
    "O",
    "B",
    "I",
    "O",
    "O",
    "O",
    "O",
    "O",
    "O",
    "O",
    "O",
    "O",
    "O"
```



```
"2": {
      "text": "Easy to start up and does not overheat as much as other laptops ."
      "labels": [
        "O",
        "O",
        "B",
        "I",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O"
      ]
    },
```