| Configuration | MSE |
|---|---|
| CNN-text | 0.53 |
| CCRF+SVR | 0.36 |
| LSTM-text | 0.35 |
| DistantSup | 0.35 |
| DeClarE (Plain) | 0.34 |
| DeClarE (Full) | **0.29** |

Table 4: Comparison of various approaches for credibility regression on NewsTrust dataset.

first three models described in Section 4.2 as baselines. For CNN-text and LSTM-text, we add a linear fully connected layer as the final layer of the model to support regression. Additionally, we also consider the state-of-the-art CCRF+SVR model based on Continuous Conditional Random Field (CCRF) and Support Vector Regression (SVR) proposed by Mukherjee and Weikum (2015). The results are shown in Table 4. We observe that DeClarE (Full) outperforms all four baselines, with a 17% decrease in MSE compared to the best-performing baselines (i.e., LSTM-text and Distant Supervision). The DeClarE (Plain) model performs substantially worse than the full model, illustrating the value of including attention and source embeddings. CNN-text performs substantially worse than the other baselines.

### 4.4 Results: SemEval

On the SemEval dataset, the objective is to perform credibility classification of a tweet while also producing a classification confidence score. We compare the following approaches and consider both variants of the SemEval task: (i) *NileTMRG* (Enayet and El-Beltagy, 2017): the best performing approach for the *close* variant of the task, (ii) *IITP* (Singh et al., 2017): the best performing approach for the *open* variant of the task, (iii) DeClare (Plain): our approach with only biLSTM (no attention and source embeddings), and (iv) DeClarE (Full): our end-to-end system with biLSTM, attention and source embeddings.

We use the evaluation measure proposed by the task's organizers: macro F1-score for overall classification and Root-Mean-Square Error (RMSE) over confidence scores. Results are shown in Table 5. We observe that DeClarE (Full) outperforms all the other approaches — thereby, re-affirming its power in harnessing external evidence.

| Configuration | Macro Accuracy | RMSE |
|---|---|---|
| IITP (Open) | 0.39 | 0.746 |
| NileTMRG (Close) | 0.54 | 0.673 |
| DeClarE (Plain) | 0.46 | 0.687 |
| DeClarE (Full) | **0.57** | **0.604** |

Table 5: Comparison of various approaches for credibility classification on SemEval dataset.

## 5 Discussion

### 5.1 Analyzing Article Representations

In order to assess how our model separates articles reporting false claims from those reporting true ones, we employ dimensionality reduction using Principal Component Analysis (PCA) to project the article representations ($g$ in Equation 2) from a high dimensional space to a 2d plane. The projections are shown in Figure 2a. We observe that DeClarE obtains clear separability between credible versus non-credible articles in Snopes dataset.

### 5.2 Analyzing Source Embeddings

Similar to the treatment of article representations, we perform an analysis with the claim and article source embeddings by employing PCA and plotting the projections. We sample a few popular news sources from Snopes and claim sources from PolitiFact. These news sources and claim sources are displayed in Figure 2b and Figure 2c, respectively. From Figure 2b we observe that DeClarE clearly separates fake news sources like *nationalreport*, *empirenews*, *huzlers*, etc. from mainstream news sources like *nytimes*, *cnn*, *wsj*, *foxnews*, *washingtonpost*, etc. Similarly, from Figure 2c we observe that DeClarE locates politicians with similar ideologies and opinions close to each other in the embedding space.

### 5.3 Analyzing Attention Weights

Attention weights help understand what DeClarE focuses on during learning and how it affects its decisions – thereby, making our model transparent to the end-users. Table 6 illustrates some interesting claims and salient words (highlighted) that DeClarE focused on during learning. Darker shades indicate higher weights given to the corresponding words. As illustrated in the table, DeClarE gives more attention to important words in the reporting article that are relevant to the claim and also