2024 IARIA Annual Congress on Frontiers in Science, Technology, Services, and Applications
IARIA Congress 2024
June 30, 2024 to July 04, 2024 - Porto, Portugal

# Automatic Assessment of Student Answers using Large Language Models Decoding Didactic Concepts

Daniel Schönle, Christoph Reich, Djaffar Ould Abdeslam

Presenter: Daniel Schönle (daniel.schoenle@hs-furtwangen.de)
Furtwangen University, Institute IDACUS
Université de Haute Alsace, IUT de Mulhouse, Institut IRIMAS



# Daniel Schönle

daniel.schoenle@hs-furtwangen.de


Research Topics

- Blockchain, Cyber-physical Systems
- Machine Learning in e-Learning Environments
- Natural Language Processing, Feature Selection
- ML-Efficiency

Publications

- Linguistic Driven Feature Selection for Text Classification as Stop Word Replacement, (2023)
- Data-Driven Tutoring: challenges and prospects, (2021)
- Industry use cases on blockchain technology, (2021)
- Digital twin as a service (2021)

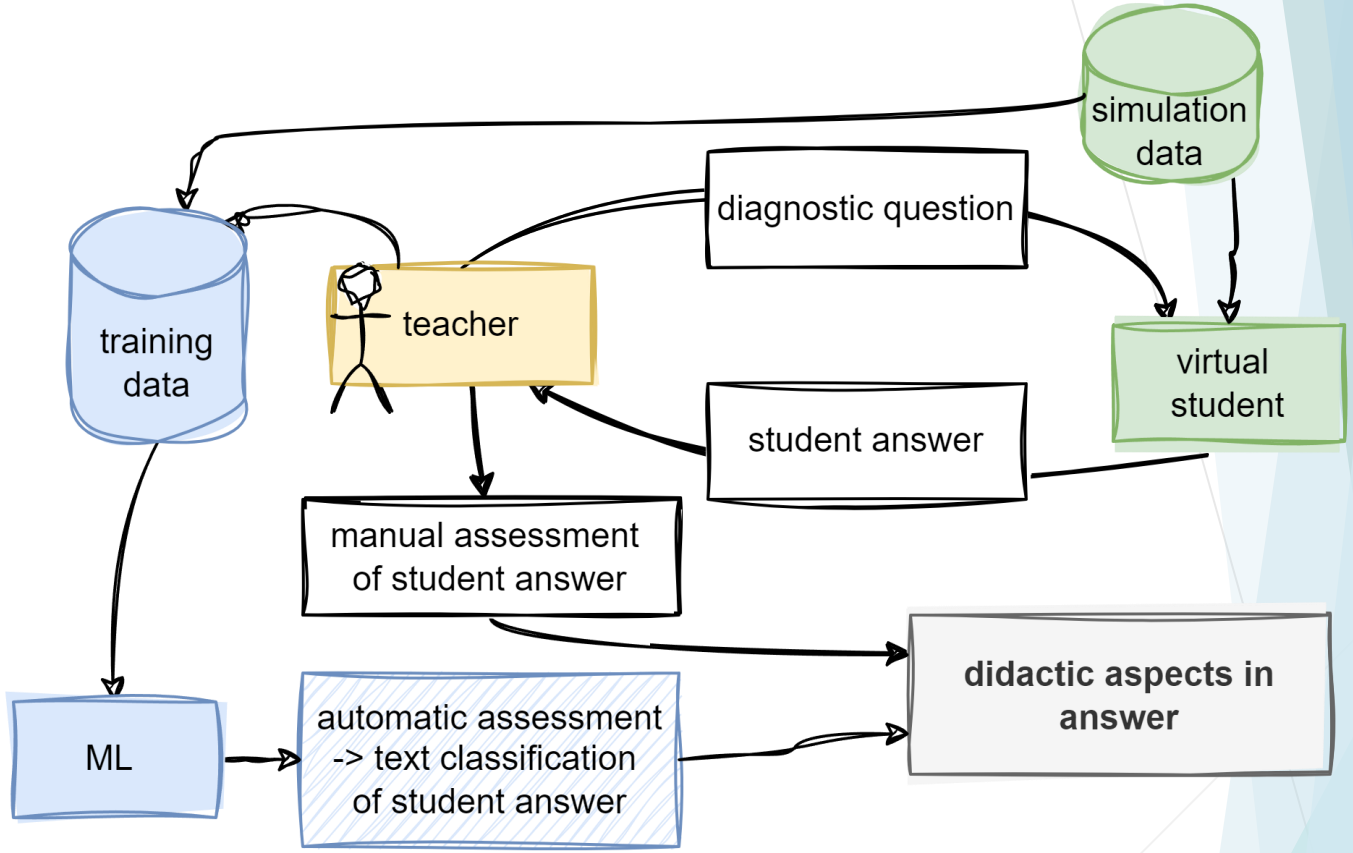
# Context

**Teaching Simulation**

- Teacher asks simulated students
- Assessment of student answer
- Open text interactions

**Assist the Teacher by ML**

- Automatic assessment of student answers
- Text classification
- Attain high-quality



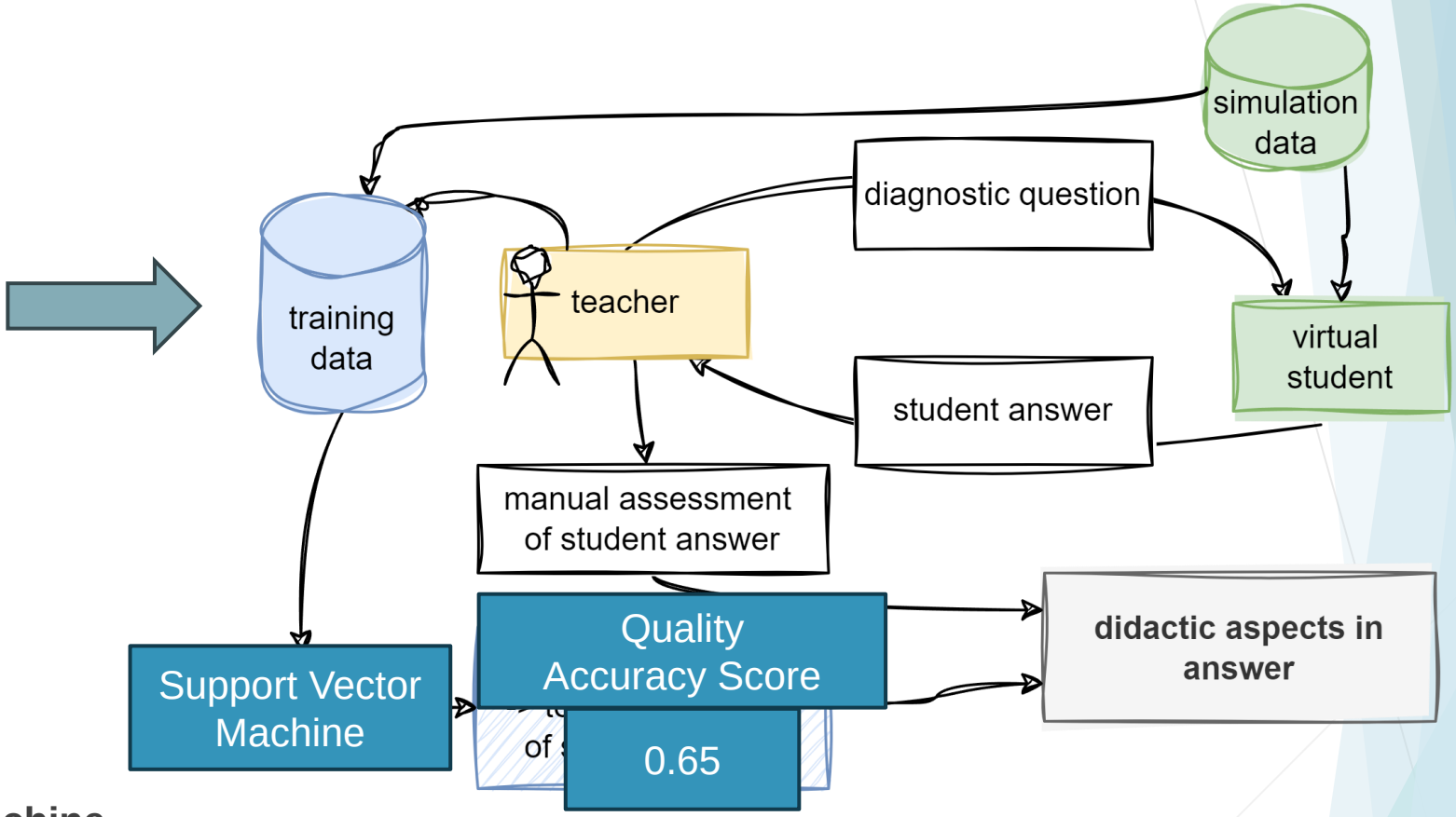
# Motivation

**Training Data**

- **Multi-Label Classification**
  9 labels to predict!
- **Data Limitation**
  350 instances,
  -> small number of samples
- **Sample Length**
  160 char mean length
  -> short text length

**Quality**

- **Required**
  High-quality prediction
  Accuracy > 0.95
- **Provided by Support Vector Machine**
  **->** Accuracy of 0.65!



# Approach

**Aim**

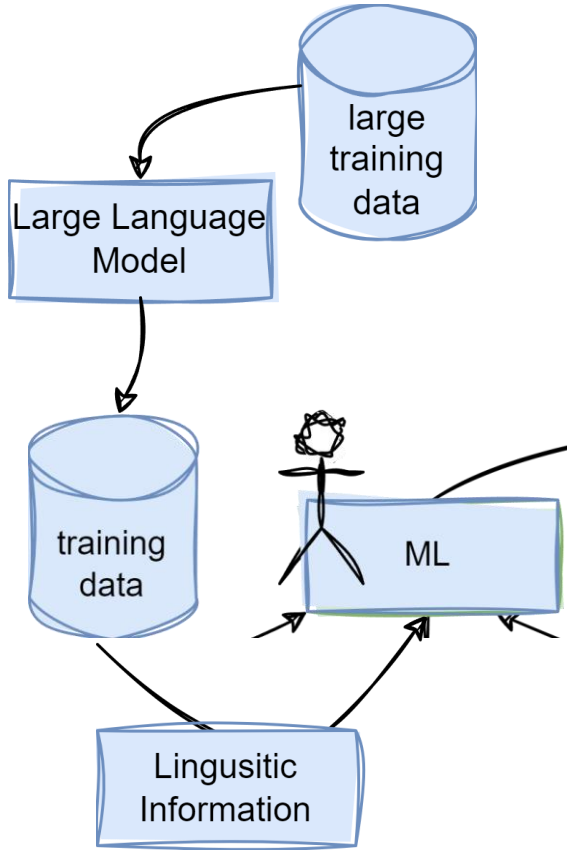
- Increase the information gain to increase text classification quality.

**Approach**

Text is language encoded information.

Increase information gain by linguistic enhancement.

- Use the linguistic skill of LLMs.
- Use linguistic context information in preprocessing.




| The | best | apples | are | sweet |
|---|---|---|---|---|
| Determiner | Superlative adjective | Plural noun | Present tense verb | Adjective |

# Information Gain by Large Language Models

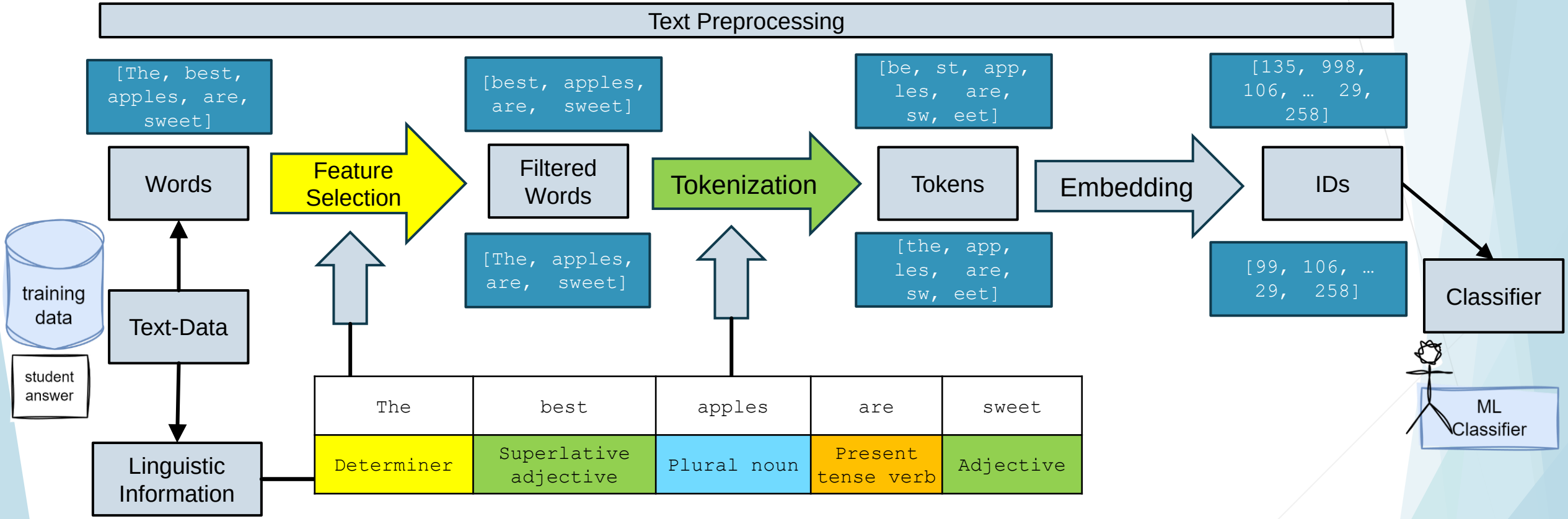
**Preprocessing: LLM-Based Data Augmentation**

- Use LLMs to modify text to increase sample count.
- Publicly available tool by DeepL SE.

**Classifier: LLM-Based Classifiers**

- Employ LLMs as foundational model.
- LLMs gain information from initial training data.
- Fine-tuning on use case data for specific classification task.

# Information Gain by Linguistic enhanced Preprocessing



# Linguistic Enhanced Feature Selection

**Objective**

- Enhance information density by removing frequent words.

**Standard Approach**

- Remove stop words.
- Example: ‘and’.

**Linguistic Enhancement**

- Weighted Unimportant Feature Selection WUP.
- Remove **word types** that are less important for the classifier model.
- Example: superlative adjectives (e.g., highest, brightest).


# Linguistic Enhanced Tokenization

**WordPiece Tokenization**

- State-of-the-Art Tokenization
- ‘Cut the text into very short pieces’

**LinPair Approach**

- Integrates linguistic information into data corpus
- ‘Cut the text into very short pieces but keep linguistic information’
- **SmartLinPair Tokenization**
  - Inject word-type tags into the dataset only when WordPiece tokenization fails.
- **CompleteLinPair Tokenization**
  - Handle word-type tags for all subtokens.

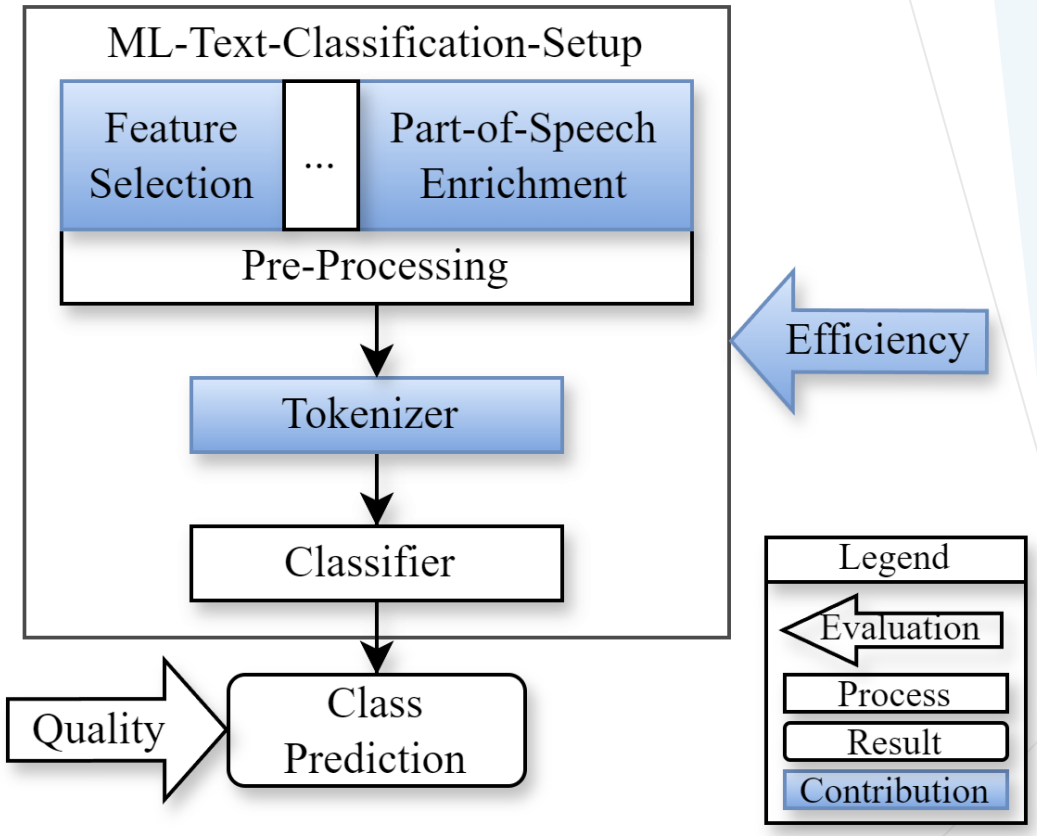
```
bugs annoi
["b", "##u, "##gs", UNKOWN]
```

```
bugs annoi
["b", "##u, "##gs", verb]
```

```
bugs annoi
["b_noun", "##u_noun",
"##gs_noun", verb]
```

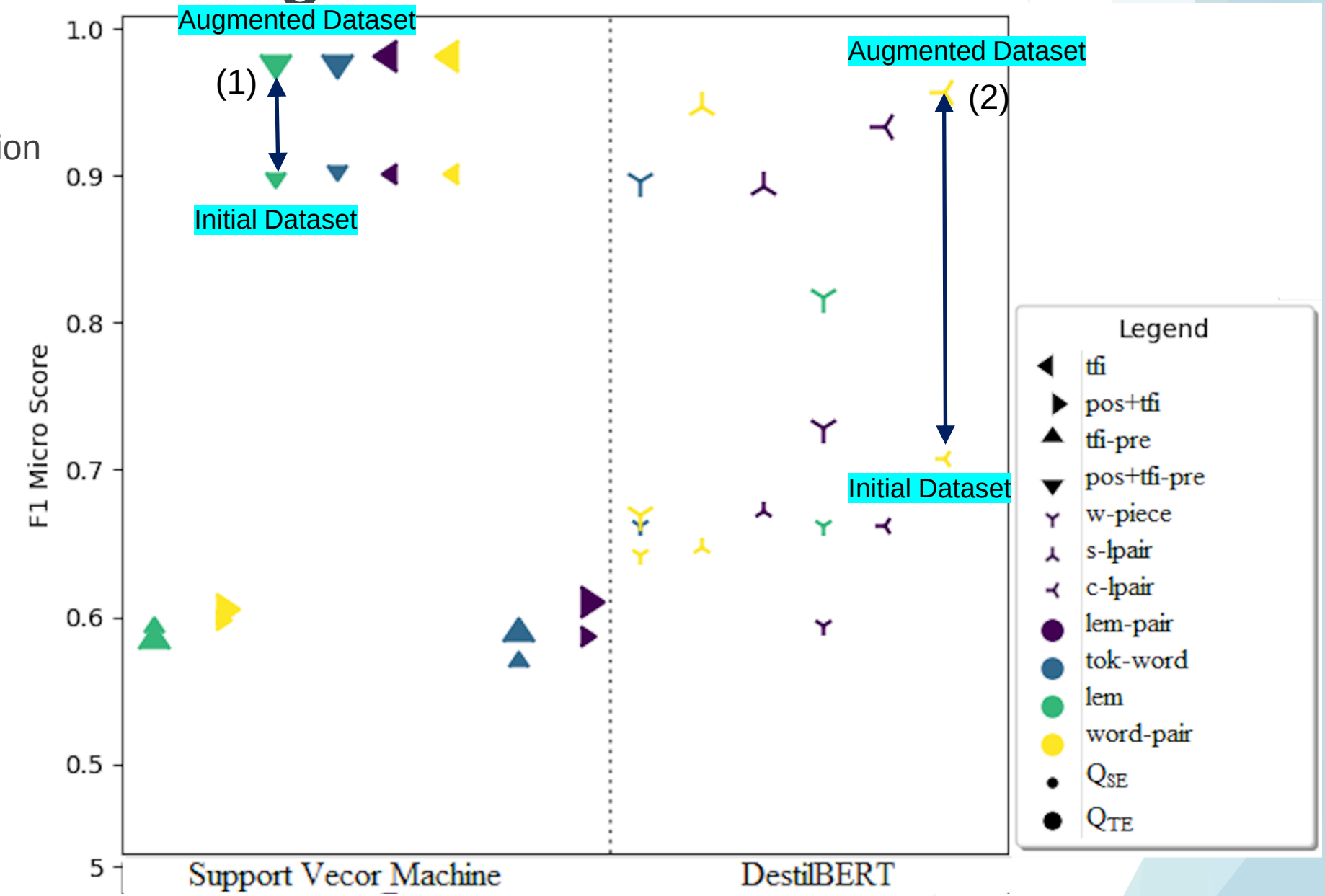
# Evaluation

- Data Augmentation
- WUP enhanced feature selection
- **Tokenization**
  - Support Vector Machine
    - Linguistic enhanced TFIDF
  - DistilBERT (Large Language Model)
    - WordPiece
    - SmartLinPair
    - CompleteLinPair
- **Dataset**
  - Labeled Student Answers
- **Metrics**
  - Quality
  - Efficiency



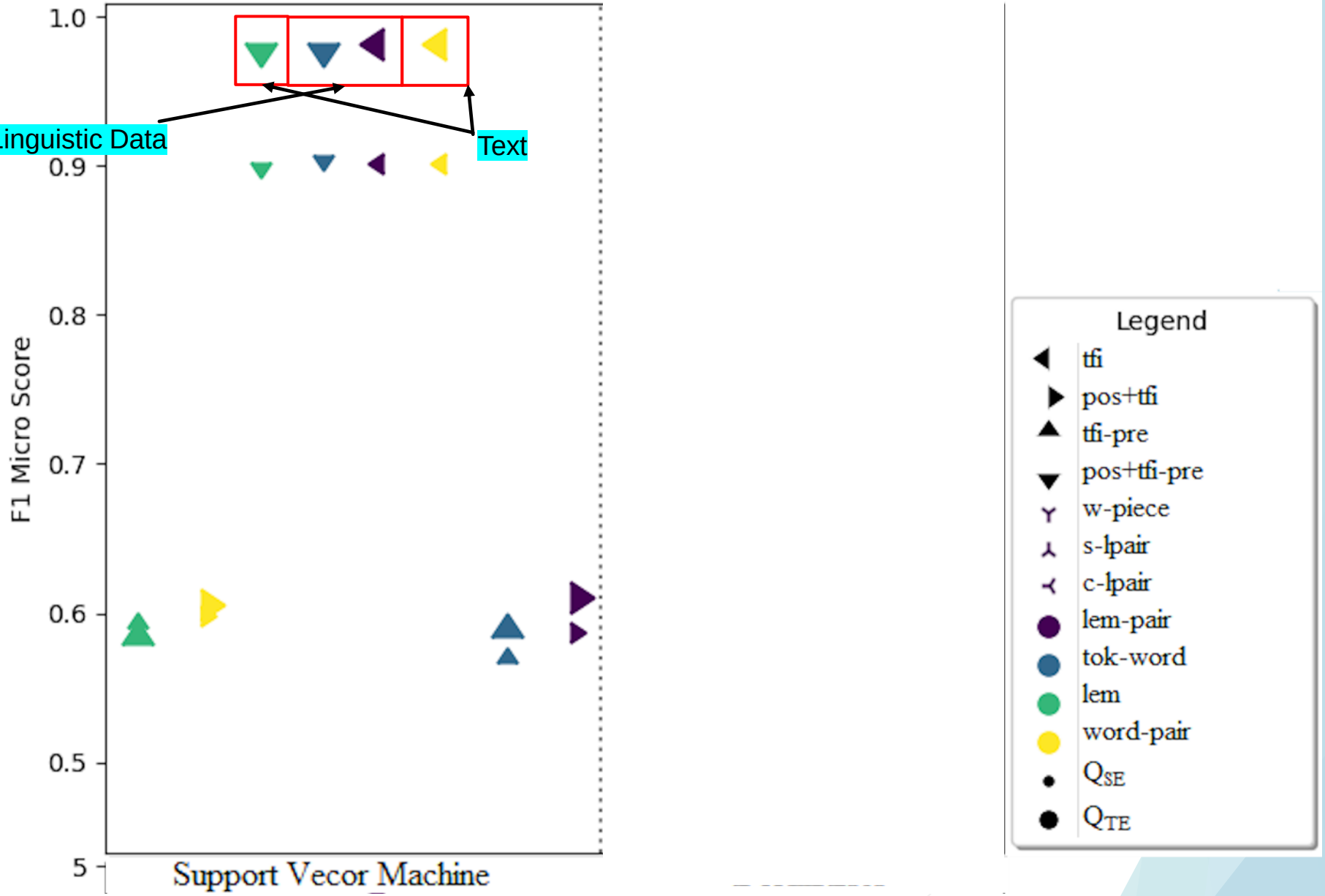
# Results LLM Data Augmentation

- All Classifiers showed increased text qualification quality by Data Augmentation
- SVM 0.96% (1)
- DestilBERT 0.95% (2)



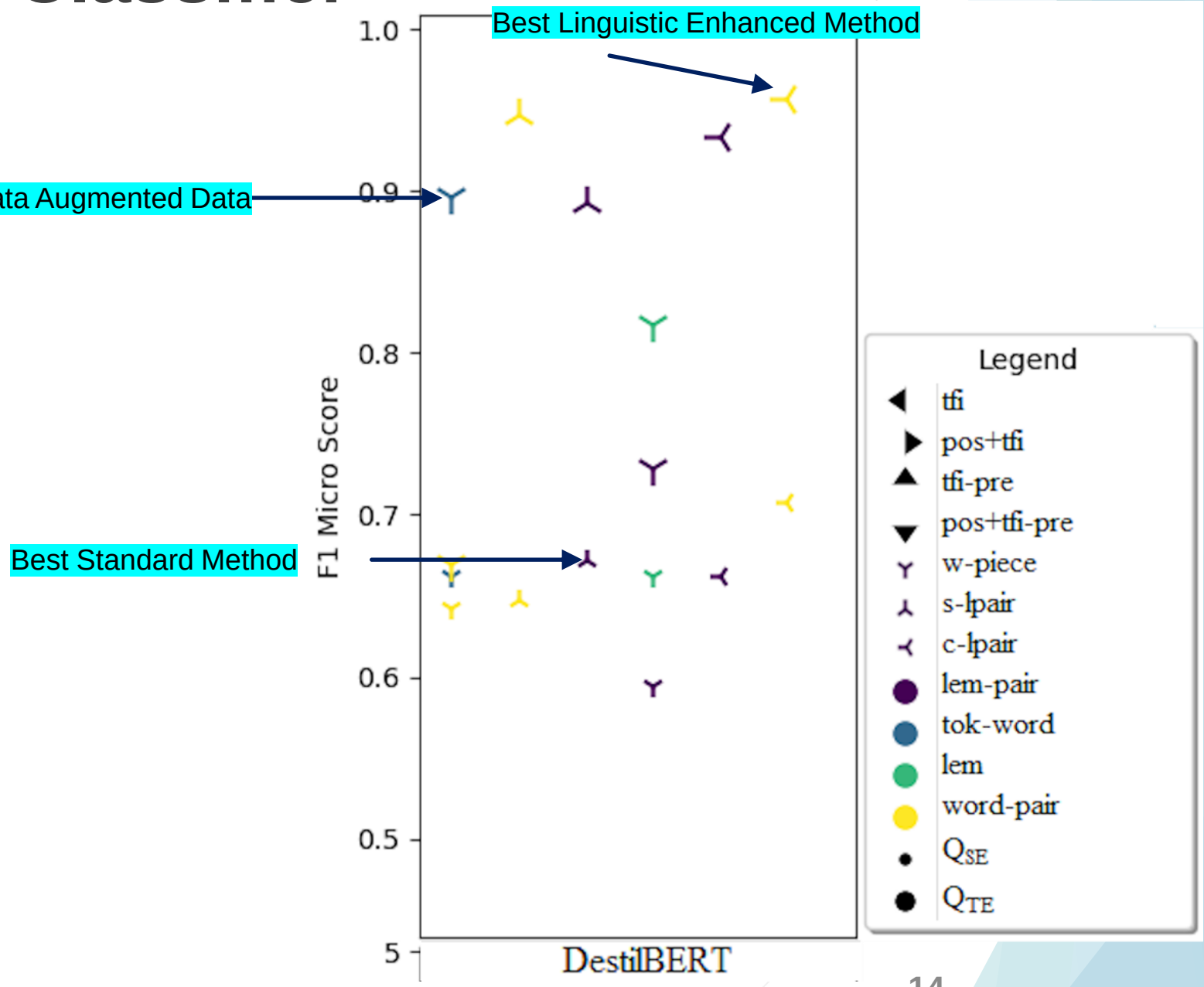
# Results SVM Classifier

- Overall High Quality by Support Vector Machine
- No positive effects by including linguistic information



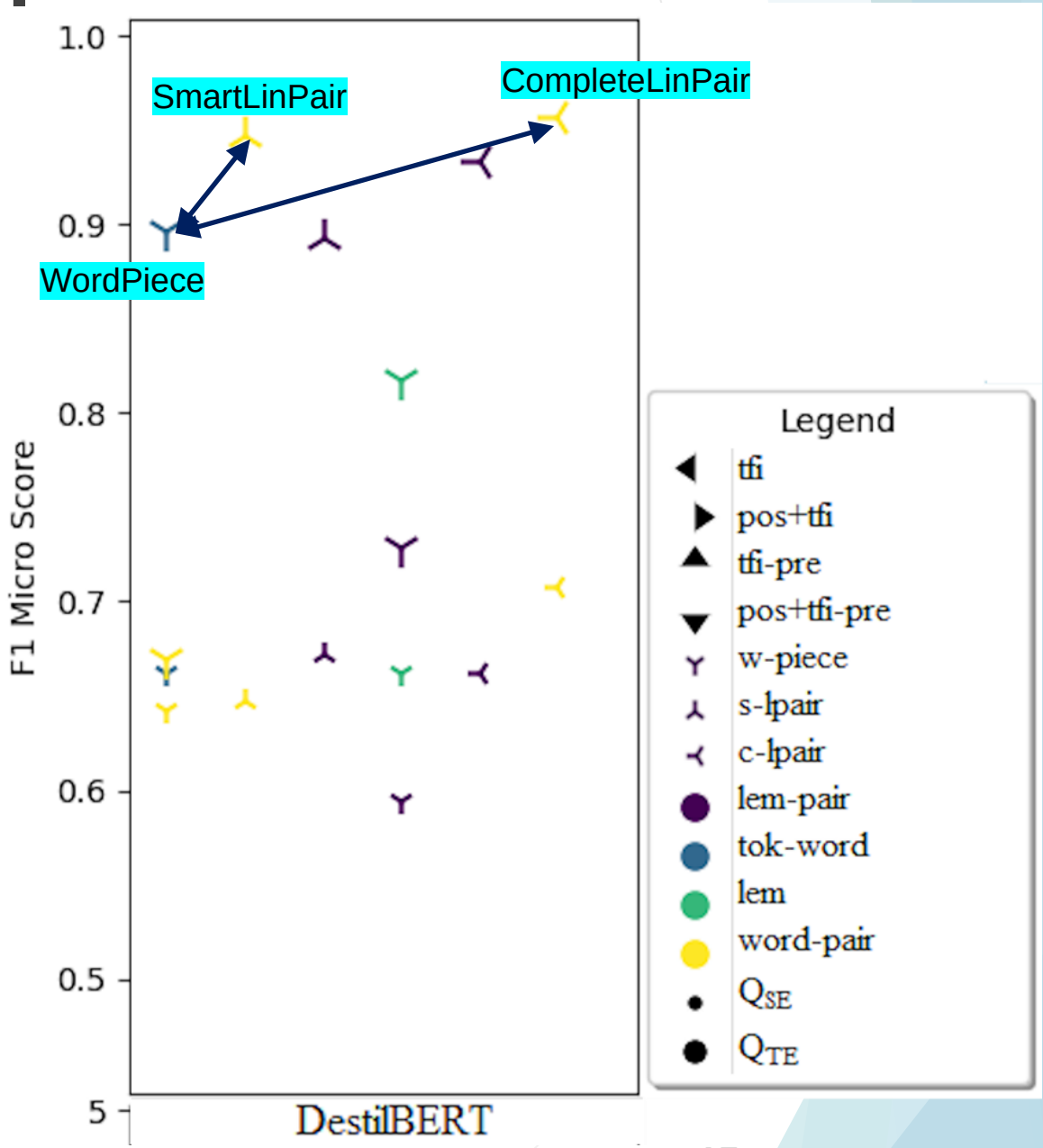
# Results DestilBERT Classifier

- The fine-tuned LLM achieved best results by using LinPair tokenization.
- Comparison by F1 micro score
- Standard Method (0.67)
- Data Augmented Data (0.89)
- Linguistic Enhanced Method (0.96)



# Results LinPair Tokenization

- Tokenization comparison by F1 micro score
- CompleteLinPair (0.96) outperform
- SmartLinPair (0.94) and
- WordPiece (0.89)



# Effects of Linguistic Enhancement

**Classifier**

- LLM DistilBERT (1) improved by 17%,
- Support Vector Machine (2) decreased.

**Tokenisation**

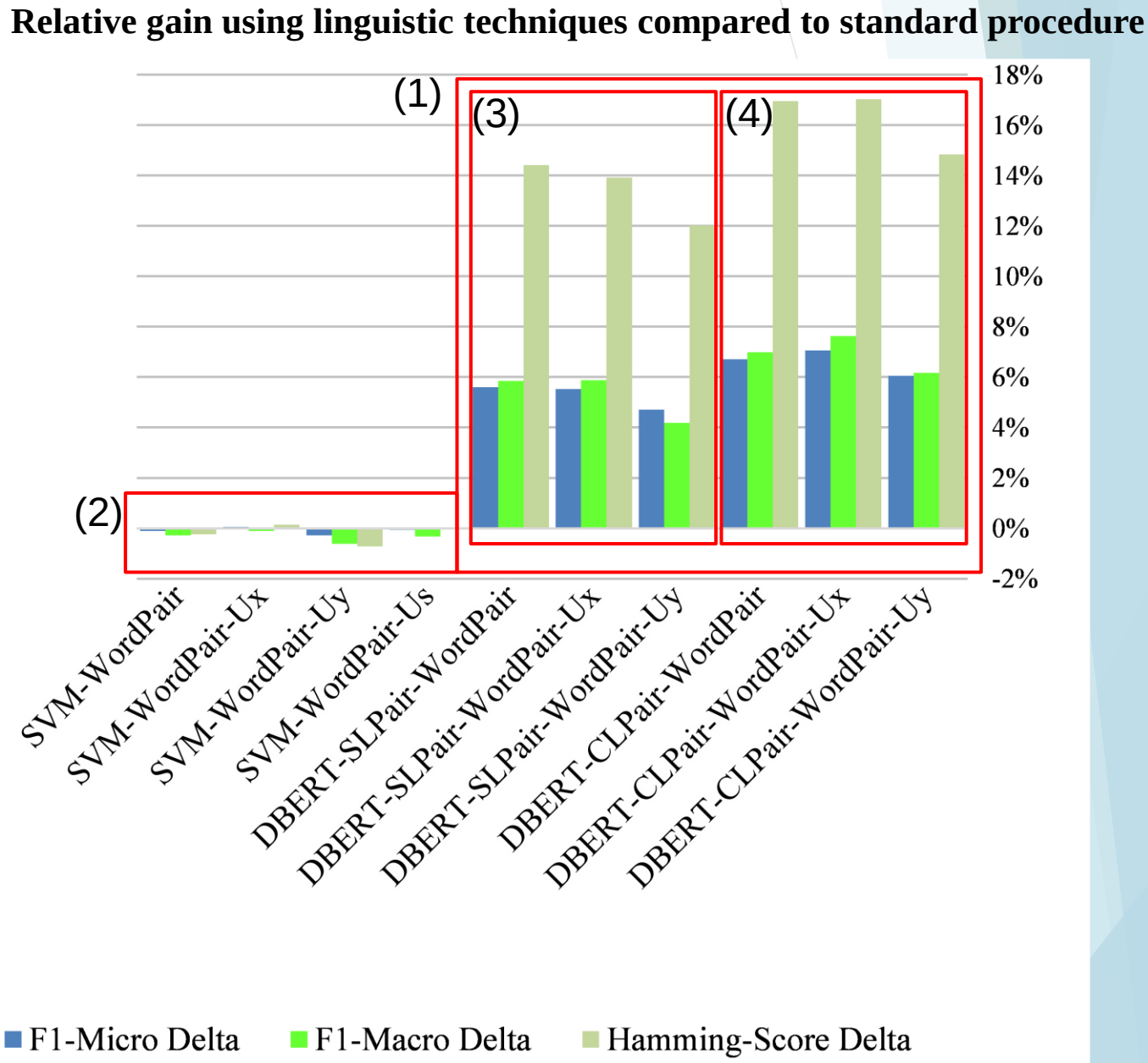
- SmartLinPair (3) and CompleteLinPair (4) leads to increase of classification quality.

**Feature Selection**

- WUP (Ux, Uy, and Us) had a negative impact.

Linguistically enhanced preprocessing yields improvements in text classification performance.

Using a finetuned LLM and CompleteLinPair the increase is 7.5% F1-Micro Score
and 16% Hamming Score.



# Thank you.


- Daniel Schönle
- schoenledanielhfu@gmail.com