



ISSN NO. 2320-5407

ISSN(O): 2320-5407 | ISSN(P): 3107-4928

# International Journal of Advanced Research

Publisher's Name: Jana Publication and Research LLP

www.journalijar.com

## REVIEWER'S REPORT

Manuscript No.: IJAR-57039

Title: Intelligent Classification of Educational Questions Using XLNet-CNN and Deep Contextual Embeddings.

### Recommendation:

Accept as it is .....

Accept after minor revision.....

**Accept after major revision .....**

Do not accept (*Reasons below*) .....

Rating	Excel.	Good	Fair	Poor
Originality			√	
Techn. Quality		√		
Clarity			√	
Significance		√		

Reviewer's ID: JPR- 212

### Detailed Reviewer's Report

**Summary of the Manuscript:** This paper proposes a hybrid deep learning model (XLNet + CNN) for automatically classifying examination questions according to Bloom's Taxonomy cognitive levels. The authors compare their proposed model with a RoBERTa + CNN baseline from prior work (Gani et al., 2023) and report comparable performance, with the advantage of higher precision on higher-order cognitive categories.

#### Overall

#### Assessment:

The manuscript addresses a practically relevant problem in educational technology—automated cognitive level classification of exam questions. The use of XLNet with CNN is technically sound, and the paper demonstrates familiarity with contemporary NLP methods. However, the manuscript suffers from **significant structural, presentational, and analytical shortcomings** that require major revision before acceptance.

#### Major Concerns Requiring Revision

##### 1. Incomplete and Corrupted Results Section (Critical Issue)

- **Pages 9–10 contain extensive corrupted/misformatted numerical data** (long sequences of "0 0 0 0") that appear to be a failed rendering of a confusion matrix or classification report.
- Figure 4 ("Classification Report") and Figure 5 ("Confusion Matrix") are referenced but **not provided**—only corrupted text remains.
- Without proper visualizations or correctly formatted results tables, the findings cannot be properly evaluated or reproduced.

##### 2. Missing Quantitative Comparison with Baseline

- The authors claim "comparable performance" to RoBERTa + CNN but provide **no explicit accuracy or F1-score values** for either model in the text.
- Table 2 is incomplete—it lists only tokenization hyperparameters, not CNN architecture details (kernel sizes, number of filters, dropout rates, etc.).
- A proper comparative table (e.g., Accuracy, Precision, Recall, F1-score for both models across all six Bloom's categories) is essential.

##### 3. Dataset Description Lacks Transparency

- The paper mentions merging five datasets, including one newly collected dataset of 1,200 questions. However:
  - Sources of existing datasets are not cited.
  - Inter-annotator agreement among educators who labeled the new dataset is not reported.
  - The class distribution (Figure 1) is referenced but the figure is missing or unclear.
- Class imbalance is acknowledged as a limitation but not addressed methodologically (e.g., oversampling, class weights).

##### 4. Structural and Organizational Problems

- Section 1 (Introduction) is overly long and includes content that belongs in Methods (e.g., detailed taxonomy descriptions).
- Section 2 (Motivation and Contribution) is very brief and could be merged with Introduction.
- The abstract claims "higher precision at higher cognitive levels," but no supporting data is presented in the results section.
- The conclusion (Section 6) correctly identifies limitations but offers generic future directions without specific next steps.

### REVIEWER'S REPORT

#### 5. Language and Technical Errors

- Typo in Section 2: "assesssavious" should be "assesses various".
- Inconsistent citation formatting: some are numbered [1], others appear as [12] with no consistent style.
- Algorithm 1 (pages 5–6) is overly detailed for a research paper; it reads like implementation notes rather than a methodological description.
- Section 5 title: "RESULT AND DISCUSSION" should be "RESULTS AND DISCUSSION" (grammar).

#### Minor Suggestions

- **Abstract:** The claim "offers performance comparable to existing models" is vague. Specify the accuracy difference (e.g., "0.5% lower overall but 3% higher precision for Evaluation category").
- **Keywords:** Add "Bloom's Taxonomy" and "Question Classification."
- **Related Work:** The section cites [12] extensively but does not critically differentiate the current work from Gani et al. (2023). Explain what novelty XLNet brings beyond RoBERTa.
- **Hyperparameter Table:** Expand Table 2 to include CNN-specific parameters (filters, kernel size, pooling strategy, dropout, learning rate, batch size, epochs).
- **Figures:** All missing figures (Figure 1: class distribution; Figure 4: classification report; Figure 5: confusion matrix; word clouds) must be provided or removed if not essential.

#### Ethical and Reproducibility Concerns

- The paper states that the dataset was "trained and evaluated exclusively on the combined dataset as given in [12]." If the dataset is identical to Gani et al. (2023), this should be explicitly acknowledged to avoid self-plagiarism or redundant publication concerns.
- Code and data availability statement is missing. For reproducibility, authors should provide a link to the dataset or code repository.

#### Recommendation for the Editor

Accept after major revision provided the authors:

1. **Repair or remove** the corrupted numerical data on pages 9–10 and provide properly formatted figures (confusion matrix, classification report, class distribution).
2. **Include explicit quantitative results** (accuracy, precision, recall, F1-score) for both XLNet+CNN and RoBERTa+CNN, ideally in a comparative table.
3. **Expand the methodology** to include full hyperparameters, dataset sources, and inter-annotator agreement.
4. **Restructure the paper** to avoid redundancy and improve readability.
5. **Correct language errors** and standardize citation formatting