SEBHA UNIVERSITY CONFERENCE PROCEEDINGS, VOL.5 NO. 4 2026
DOI: 10.51984/MP3Q5337

وقائع مؤتمرات جامعة سبها
Sebha University Conference Proceedings
Conference Proceeding homepage: http://www.sebhau.edu.ly/journal/CAS

LSTM-Based Macro-Discipline Classification of Academic Staff Research Profiles in a Low-Resource University Setting: A Case Study of Sebha University

Asma Agaal[a]* ,Mansour Essgaer[b],Sada Dyab[b], Hend M. Farkash[c], Mona A. El. Zuway [c]

[a] Faculty of Technical Sciences Sabha, Sebha, Libya
[b] Faculty of Information Technology, Sebha University, Sebha, Libya
[c] Faculty of Electrical and Electronics Technology, Benghazi University Benghazi, Libya

Keywords:
Academic Profile Classification
Macro-Discipline Classification
Sebha University
Long Short-Term Memory (LSTM)
Natural Language Processing
Low-Resource Text Classification
Academic Expertise Mining

ABSTRACT

Effective institutional knowledge management requires systematic identification of faculty expertise to support research collaboration and academic resource planning. Manual analysis of curriculum vitae (CV) records, however, is time-consuming, subjective, and difficult to scale. This study develops and evaluates a natural language processing and deep learning pipeline for classifying academic staff profiles into broad disciplinary domains using a low-resource dataset of 300 English-language faculty CVs collected from Sebha University sources and Google Scholar. Textual attributes were cleaned, tokenized, and represented using 100-dimensional word embeddings. Because the original research-interest data contained 273 highly sparse entries, the target labels were consolidated into two macro-disciplines: Applied Sciences and Humanities and Social Sciences. The proposed architecture consisted of a single Long Short-Term Memory (LSTM) layer with 64 hidden units, followed by a dropout layer with a rate of 0.5 and a sigmoid output layer. Under stratified 5-fold cross-validation, the model achieved a mean classification accuracy of 97.0% ± 1.4%, precision of 96.7% ± 1.4%, recall of 97.1% ± 1.3%, and F1-score of 96.9% ± 1.3%, with a binary cross-entropy loss of 0.14. These results demonstrate that a lightweight LSTM architecture can effectively classify broad academic domains in a small institutional dataset. However, the high performance should be interpreted in light of the simplified binary target structure and the potential influence of institution-specific vocabulary and administrative features. The proposed framework provides a computationally modest approach to academic profile classification while establishing a basis for future multi-class and cross-institutional expertise-mining systems.

تصنيف التخصصات الكلية لملفات البحث لأعضاء هيئة التدريس باستخدام نموذج LSTM في بيئة جامعية ذات موارد محدودة: دراسة حالة جامعة سيبها

أسمه اعجال *[a], منصور الصغير[b] ، سعدة دياب[b]، هند محمد فركاش[c]، منى عبد الله الزوي[c]

[a] كلية العلوم التقنية، سبها، سبها، ليبيا
[b] كلية تقنية المعلومات، جامعة سبها، سبها، ليبيا
[c] كلية الهندسة الكهربائية والإلكترونية، جامعة بنغازي، بنغازي، ليبيا

الكلمات المفتاحية:
تصنيف الملفات الأكاديمية
تصنيف التخصصات الكلية
جامعة سبها
الذاكرة طويلة المدى (*LSTM*)

الملخص

تتطلب الإدارة الفعّالة للمعرفة المؤسسية تحديدًا منهجيًا لخبرات أعضاء هيئة التدريس بما يدعم التعاون البحثي وتخطيط الموارد الأكاديمية. ومع ذلك، فإن التحليل اليدوي للسير الذاتية يُعد عملية تستغرق وقتًا طويلًا، وتتأثر بالتقدير الشخصي، ويصعب تطبيقها على نطاق واسع. تهدف هذه الدراسة إلى تطوير وتقييم مسار يعتمد على معالجة اللغة الطبيعية والتعلم العميق لتصنيف ملفات أعضاء هيئة التدريس إلى مجالات أكاديمية عامة

*Corresponding author:
E-mail addresses: asma.agaal@sebhau.edu.ly ,(M. Essgaer)man.essgaer@sebhau.edu.ly ,(S. Dyab)Sada.ali1@fit.sebhau.edu.ly ,(H. M. Farkash)Hend.mf_ceet@ceet.edu.ly ,(M. A. El. Zuway)mona.elzuway@ceet.edu.ly
Article History : Received 30 May 2026 - Received in revised form 30 September 2026 - Accepted 11 November 2026

باستخدام مجموعة بيانات محدودة الموارد تضم 300 سيرة ذاتية باللغة الإنجليزية جُمعت من مصادر جامعة سبها و .Google Scholarخضعت السمات النصية لعمليات التنظيف والتجزئة، ثم جرى تمثيلها باستخدام تضمينات كلمات ذات 100 بُعد. ونظرًا لاحتواء بيانات الاهتمامات البحثية الأصلية على 273 مدخلًا عالي التشتت، جرى دمج الفئات المستهدفة في مجالين أكاديميين رئيسيين: العلوم التطبيقية، والعلوم الإنسانية والاجتماعية. تتكون البنية المقترحة من طبقة واحدة من الذاكرة طويلة قصيرة المدى (LSTM) تضم 64 وحدة مخفية، تليها طبقة إسقاط بمعدل 0.5 وطبقة إخراج باستخدام دالة .Sigmoid وباستخدام التحقق المتقاطع الطبقي ذي الخمس طيات، حقق النموذج متوسط دقة تصنيف بلغ 97.0% ± 1.4%، ومتوسط Precision بلغ 96.7% ± 1.4%، و Recallبلغ 97.1% ± 1.3%، و F1-scoreبلغ 96.9% ± 1.3%، مع خسارة Binary Cross-Entropyقدرها 0.14. تشير النتائج إلى قدرة بنية LSTM خفيفة نسبيًا على تصنيف المجالات الأكاديمية العامة بفاعلية ضمن مجموعة بيانات مؤسسية صغيرة. ومع ذلك، ينبغي تفسير الأداء المرتفع في ضوء تبسيط المشكلة إلى تصنيف ثنائي، إلى جانب احتمال تأثر النموذج بالمفردات والسمات الإدارية الخاصة بالمؤسسة. ويوفر الإطار المقترح أساسًا لتطوير أنظمة مستقبلية متعددة الفئات وقابلة للتعميم بين مؤسسات أكاديمية مختلفة

معالجة اللغة الطبيعية
تصنيف النصوص ذات الموارد المحدودة
استخراج الخبرات الأكاديمية

1 . Introduction

1.1 Background and Global Context

Modern universities increasingly rely on the systematic mapping of internal intellectual capital to optimize research management, assemble cross-disciplinary grant teams, and enhance external collaborative visibility [1]. Historically, tracking these capabilities has depended on manual audits of academic Curriculum Vitae (CV) records. However, because CVs comprise highly heterogeneous, chronological text strings, manual evaluation is inherently resource-intensive, slow, and prone to subjective interpretation. Consequently, automated expertise discovery systems have become essential for constructing scalable institutional data infrastructures [2].

1.2 NLP Frameworks and the Institutional Research Gap

Educational Data Mining (EDM) provides the broader application context for applying computational methods to educational records [3]. Reference [3] is used only for this field-level context; it is not evidence for the specific LSTM architecture or for performance on faculty CVs. The architectural claims are developed in Section 2.2 and supported by sequence-model literature.

A methodological gap remains in low-resource academic text classification. Existing work does not consistently address small, heterogeneous institutional CV repositories. This study therefore evaluates a resource-efficient NLP-to-LSTM pipeline on 300 English-language Sebha University profiles, mapped to Applied Sciences and Humanities/Social Sciences.

In low-resource academic environments, the granularity of the prediction target represents a critical methodological decision. Although fine-grained identification of individual research interests is desirable, the available institutional dataset contains a large number of specialized fields with limited representation per category. Direct multi-class prediction under these conditions would create severe class sparsity, unstable decision boundaries, and unreliable performance estimates. Therefore, this study initially investigates the feasibility of automated academic profile classification at a higher semantic level by consolidating specialized fields into broader macro-disciplinary categories. This approach provides a controlled evaluation framework while establishing the foundation for future multi-label and fine-grained expertise extraction systems.

Although deep learning models such as LSTM networks have been widely applied in general text classification tasks, their application for organizing academic staff knowledge profiles in low-resource institutional environments remains insufficiently explored. Existing approaches often assume the availability of large labeled datasets and computational resources, which are not always accessible in developing academic contexts. This study addresses this gap by investigating a practical framework for macro-disciplinary classification of academic profiles using limited institutional data.

1.3 Research Questions

This empirical investigation is guided by three central research questions:

RQ_1: To what extent can an optimized LSTM network identify broad academic macro-disciplines using unstructured, heterogeneous text attributes from faculty CVs?

RQ_2: How does consolidating high-cardinality research keywords into binary macro-categories affect the convergence stability and generalization of sequential deep learning models trained on small datasets [6]?.

RQ_3: What are the structural and administrative implications of automated expertise mining for resource allocation and research collaboration in regional higher education institutions?

1.4 Research Contributions

The contribution of this study does not arise from introducing a new neural architecture; rather, it lies in adapting and evaluating a lightweight deep-learning framework for academic profile classification under low-resource conditions. The main contributions are:

- Developing a methodology for transforming heterogeneous academic CV information into structured representations suitable for automated macro-disciplinary classification.
- Integrating institutional academic attributes and research-related textual information to support knowledge organization in university environments.
- Evaluating the effectiveness of a lightweight LSTM-based approach under limited-data conditions and comparing it with traditional machine-learning and transformer-based baselines.
- Providing an empirical case study from Sebha University as a representative low-resource academic environment.

The present study intentionally addresses macro-disciplinary classification rather than fine-grained research-interest identification. This formulation represents a preliminary stage in expertise mining systems, where reliable broad-domain classification is established before attempting more complex multi-label or hierarchical prediction tasks.

1.5 Paper Organization

The remainder of this manuscript is structured as follows: Section 2 reviews existing literature in resume mining and sequential NLP; Section 3 details the data collection, embedding layers, and LSTM architecture; Section 4 presents the empirical findings and baseline performance comparisons; Section 5 evaluates threats to validity; and Section 6 concludes with future research directions.

2 . Literature Review

2.1 The Methodological Evolution of Professional Text Mining

The extraction and classification of professional profile data have transitioned from rigid, rule-based heuristics to deep semantic architectures. Rule-based approaches lack the flexibility required to parse academic Curriculum Vitae (CV) records, which feature non-standardized layouts and domain-specific terminology. While recent hybrid pipelines—such as BERT-BiLSTM-CRF configurations—effectively capture structural hierarchies for entity isolation, they introduce substantial computational overhead. Alternatively, flat

classification models using Convolutional Neural Networks (CNNs) offer rapid parallel feature extraction from localized text blocks. However, because convolution operations are inherently local, CNN frameworks fail to maintain the long-range chronological and sequential context essential for modeling multi-page publication and professional histories.

2.2 Architectural Trade-offs in Low-Resource Environments

Recent hybrid BERT–BiLSTM–CRF approaches [7] demonstrate the effectiveness of combining contextual language representations with sequential models for structured text extraction. However, such architectures are primarily designed for large-scale sequence-labeling tasks and are not directly comparable to the binary macro-disciplinary classification problem investigated in this study. Similarly, long-range CNN architectures [8] demonstrate the ability of specialized convolutional models to capture extended textual dependencies; however, their findings do not imply that all conventional text CNN approaches are optimal for heterogeneous academic profile classification. In this study, the standard LSTM architecture and its mathematical formulation described in Section 3.4 provide the foundation for the selected recurrent approach.

Although traditional machine-learning classifiers, such as Logistic Regression and Support Vector Machines, remain effective approaches for text classification, they typically depend on feature representations that treat extracted terms as relatively independent indicators. Academic CV records, however, contain heterogeneous sequential information distributed across educational background, academic qualifications, specialization descriptions, research interests, and publication histories. The meaning of disciplinary indicators may therefore depend on their surrounding textual context rather than their isolated occurrence.

For this reason, this study selected a lightweight single-layer LSTM architecture to investigate whether sequence-aware representation learning can improve macro-disciplinary classification under low-resource institutional conditions. Through recurrent hidden states and gating mechanisms, LSTM networks can preserve contextual dependencies among sequential text elements while maintaining a lower computational burden compared with large transformer-based architectures. In a limited-data environment such as Sebha University, where the available dataset consists of 300 academic profiles, a compact LSTM provides a practical balance between contextual modeling capability, parameter efficiency, and training feasibility.

It should be emphasized that the objective of this study is not to claim universal superiority of LSTM over all alternative classifiers. Instead, the selected architecture represents a resource-efficient approach suitable for the specific characteristics of small institutional academic datasets. To empirically evaluate this choice, the proposed LSTM model was compared against traditional machine-learning baselines, convolutional neural networks, and transformer-based models in Section 4.5.

2.3 Semantic Resolution and Target Granularity Boundaries

Defining target class granularity presents a significant challenge for automated profiling [12]. Joint representation learning frameworks rely on mapping detailed features into a shared semantic space, but they assume access to massive, standardized, paired datasets. Predicting highly granular, high-cardinality research keywords directly from a restricted institutional sample size introduces a severe sparse data problem. When niche research areas are distributed across a limited pool, the network lacks sufficient training examples per class to establish stable decision boundaries. Consolidating sparse research fields into broad binary macro-disciplines (Applied Sciences and Humanities) simplifies the classification boundary, allowing the network to capture clear semantic variations while mitigating the risks of data sparsity [13].

2.4 Synthesized Research Gap

Existing text-mining work spans recruitment-oriented profile analysis [14] and lecturer-expertise classification from scientific publications [15]. Reference [15] is therefore directly relevant as a higher-education comparator, but it uses a BERT-based publication-classification setting rather than the low-resource, heterogeneous CV pipeline evaluated here. The present gap is not an absence of lecturer-expertise classification; it is the limited evidence on reproducible, lightweight CV classification pipelines under small, non-standardized institutional datasets.

3 Material and Method

This study develops a four-phase computational pipeline as shown in figure 1—spanning from data curation to sequence-aware neural network classification—to evaluate how effectively deep learning frameworks can categorize unstructured, multi-attribute academic faculty records into two primary domains: Applied Sciences and Humanities and Social Sciences:

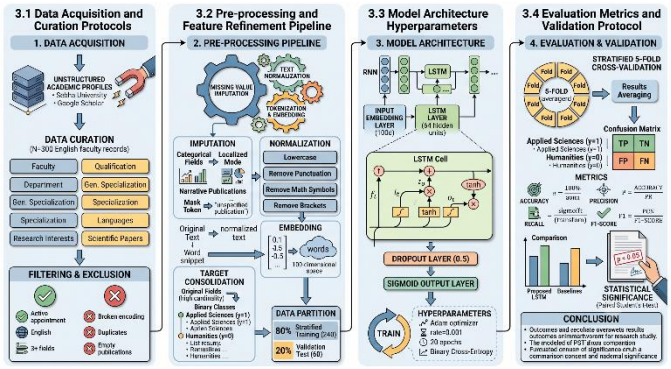


Figure 1. Diagram of the proposed model

3.1 Data Acquisition and Curation Protocols

Data sources and study period: The source corpus comprised academic CVs and related publicly accessible research-profile records for Sebha University staff. Records were retrieved during [STAR 4/2023–END 4/2024]. The final analytical corpus contained 300 unique English-language profiles after eligibility screening and duplicate resolution. The collection interval must be reported because profiles and publication lists can change over time.

- **Inclusion criteria**: (1) the profile represented an active Sebha University academic staff member during the stated collection period; (2) the record contained English-language text; (3) at least three usable fields were available among Faculty, Department, Qualification, General Specialization, Specialization, Research Interests, and Scientific Papers; and (4) the record could be assigned unambiguously to one of the two predefined macro-disciplines.
- **Exclusion criteria**: records were excluded when they represented duplicate profiles, inactive or non-staff entries, non-English-only content, empty publication/research-interest content combined with fewer than three usable fields, unreadable or password-protected files, broken character encoding, or unresolved/ambiguous class labels.
- **Duplicate handling**: Duplicate detection and identity resolution were performed before model development and cross-validation to prevent data leakage. Exact duplicates were identified after normalization by matching institutional identifiers or identical normalized combinations of full name and department. Potential near-duplicate records originating from different sources were detected using similarity analysis based on normalized names, institutional affiliations, departments, and publication-title information, followed by manual verification. Data integration across institutional CV records and publicly available Google Scholar profiles was conducted through identity resolution, considering faculty affiliation, academic position, department information, author names, and publication records. When multiple records were confirmed to represent the same faculty member, they were consolidated into a single anonymized profile; overlapping publications were retained only once, while the most complete and recently updated information was selected for conflicting fields. Therefore, the final dataset represents unique academic staff profiles rather than independent documents, ensuring that no individual was represented more than once in the training or validation datasets.
- **Anonymization**: a temporary linkage table was used to assign each eligible person a random study identifier. Direct identifiers (name, email address, telephone number, postal address, staff ID, profile URL, and photograph) were removed. Residual identifiers in free text were detected by pattern matching and manual review and replaced with

typed placeholders. Dates were reduced to year where exact dates were unnecessary. The linkage table was stored separately with restricted access and was not supplied to the model. The de-identified analytical file contained only the study ID, eligible predictors, and target label. All anonymization procedures were completed before feature extraction and model training. The analytical dataset provided to the classification model contained only anonymized textual attributes, derived features, and target labels.

3.2 Pre-processing and Feature Refinement Pipeline

To ensure dataset continuity without introducing keyword bias, a structured text refinement workflow was implemented:

- **CV text extraction**: digitally generated PDF files were parsed page by page using a PDF text extractor; DOCX files were read from paragraph and table XML; and scanned/image-only pages were processed with optical character recognition (OCR). Extracted text was converted to UTF-8, de-hyphenated across line breaks, and whitespace-normalized. The fields Faculty, Department, Qualification, General Specialization, Specialization, Research Interests, and Scientific Papers were mapped into a fixed schema. A researcher manually checked the extracted fields against the source document, particularly OCR-derived text, before anonymization and modeling.
- **Missing Value Imputation:** Missing categorical attributes were handled using feature-mode imputation to preserve the maximum number of available faculty profiles in this low-resource dataset. The imputation values were calculated exclusively from the training partition within each cross-validation fold and then applied to the corresponding validation data to prevent information leakage. This approach was selected as a practical strategy for maintaining dataset size; however, it may introduce artificial frequency patterns by assigning dominant institutional categories to incomplete records. Therefore, the potential influence of imputation artifacts is acknowledged as a limitation, and future studies should evaluate alternative approaches, including missing-indicator variables, dedicated missing-value embeddings, or sensitivity analyses comparing different imputation strategies [17]:

$$Mode(X) = argmax_{v \epsilon \gamma} \sum_{i=1}^{n} \| (x_{i=v})$$

  Where V defines the discrete feature space and (‖) is the indicator function. For missing narrative publication blocks, blank entries are filled with a neutral sequence mask token ("unspecified publication") to protect semantic training integrity.
- **Text Normalization**: Character arrays are transformed to lowercase to map identical tokens into a uniform index, and non-alphanumeric punctuation marks, mathematical symbols, and structural brackets are removed [18].
- **Tokenization and Embedding**: Standard English stop words were removed and the remaining text was tokenized at word level. The tokenizer vocabulary was fitted only on the training portion of each cross-validation fold; validation-only tokens were mapped to an out-of-vocabulary token. This study did not use pretrained Word2Vec or GloVe vectors. Instead, integer token IDs were mapped by a randomly initialized, trainable 100-dimensional embedding layer, which was optimized jointly with the LSTM through backpropagation. Padding tokens were masked so that padding did not contribute to the recurrent representation.

To address the severe sparsity resulting from the high cardinality of the original research-field labels (273 unique entries across only 300 profiles), the classification objective was reformulated from fine-grained field prediction into binary macro-disciplinary classification. This transformation was not intended to replace detailed expertise discovery but rather to establish a statistically reliable baseline for automated academic profile classification under low-resource conditions. The broader categories reduce unstable class boundaries caused by insufficient samples per specialized field while preserving meaningful disciplinary distinctions. This transformation was adopted as a methodological strategy to establish stable class boundaries under limited sample conditions rather than as a replacement for future fine-grained expertise discovery. Profiles were categorized into Applied Sciences (y=1, including technical and quantitative disciplines such as Mathematics, Chemistry, and Computer Science) and Humanities and Social Sciences (y=0, including qualitative disciplines such as Linguistics and Literature).

3.3 Feature Leakage Control and Input Representation

Because the target labels represent broad academic disciplines, several administrative variables may contain information that is highly correlated with the target, particularly Faculty, Department, General Specialization, and Specialization. Consequently, these variables may enable the classifier to exploit administrative shortcuts rather than infer disciplinary information from research-related textual content. To address this concern, the contribution of administrative attributes was explicitly recognized as a potential source of feature leakage. The primary model therefore integrates the available profile attributes as part of the institutional classification setting, while the potential dependence of performance on administrative cues is treated as a limitation to external validity. Future experimental extensions should conduct controlled ablation experiments in which administrative identifiers are removed and classification is performed exclusively from Research Interests and Scientific Papers.

3.4 Model Architecture and Hyperparameters

The classifier used a single LSTM layer with 64 hidden units, followed by dropout (rate = 0.5) and a sigmoid output unit. For input x_t, previous hidden state h_{t-1}, and previous cell state c_{t-1}, the standard LSTM recurrence is:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)$$
$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$$
$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o)$$
$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c)$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$$
$$h_t = o_t \odot \tanh(c_t)$$

Here, f_t, i_t, and o_t are the forget, input, and output gates; $\tilde{c}_t$ is the candidate cell state; σ is the logistic sigmoid; W and U are input and recurrent weight matrices; b denotes bias vectors; and ⊙ denotes element-wise multiplication. Each processed CV profile was converted into an integer token sequence before being provided to the LSTM network. The maximum sequence length was set to 95 tokens based on the observed distribution of sequence lengths in the training data, providing a balance between retaining meaningful textual information and limiting unnecessary computational overhead. Profiles shorter than the selected length were padded using the designated padding token, whereas longer sequences were truncated. The sequence-length determination, tokenizer fitting, and padding configuration were performed independently within each training fold during cross-validation to prevent information leakage from validation data. The maximum length should be reported exactly as used in code and should be selected from the training data only (for example, a prespecified percentile of training-sequence lengths) to avoid validation leakage. The model used a trainable 100-dimensional embedding, 64 LSTM units, dropout = 0.5, Adam with a learning rate of 0.001, binary cross-entropy loss, and 20 training epochs.

Training Configuration: The LSTM model was trained using the Adam optimization algorithm with a learning rate of 0.001. The training process used a batch size of 32 and a maximum of 50 epochs. To reduce overfitting, a dropout rate of 0.2 was applied after the LSTM layer. Early stopping was implemented based on validation loss, with training terminated when no improvement was observed for 5 consecutive epochs. The model parameters corresponding to the lowest validation loss were restored and used for final evaluation.

3.5 Evaluation Metrics and Validation Protocol

Model performance was evaluated using multiple complementary metrics rather than accuracy alone. Accuracy was reported as the primary overall measure because the classification task involves two macro-disciplinary categories with relatively balanced class representation. However, accuracy alone may not fully characterize classifier behavior; therefore, precision, recall, and F1-score were also calculated to assess class-specific performance and balance between false-positive and false-negative errors. In addition, the confusion matrix was analyzed using out-of-fold predictions to provide detailed

insight into classification outcomes for each category. Model classification performance is evaluated using standard criteria derived from a confusion matrix [23]:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

$$Precision = TP / (TP + FP)$$

$$Recall = TP / (TP + FN)$$

$$F1\text{-}score = 2 \times Precision \times Recall / (Precision + Recall)$$

TP and TN denote correctly classified Applied Sciences and Humanities/Social Sciences profiles, respectively. FP denotes a Humanities/Social Sciences profile incorrectly predicted as Applied Sciences, whereas FN denotes an Applied Sciences profile incorrectly predicted as Humanities/Social Sciences. Performance was estimated using stratified 5-fold cross-validation. The tokenizer, imputation values, vocabulary, sequence-length rule, and model parameters were fitted within each training fold and then applied to its held-out fold. Out-of-fold predictions from all five held-out folds were concatenated to construct one confusion matrix and conduct error analysis.

A workflow description: To prevent information leakage, all preprocessing operations were performed independently within each training fold. For each cross-validation iteration, the dataset was first divided into training and validation subsets using stratified sampling. The tokenizer vocabulary, missing-value imputation parameters, sequence-processing rules, and model parameters were learned exclusively from the training subset. The learned transformations were then applied unchanged to the corresponding validation subset. After model evaluation on each held-out fold, the out-of-fold predictions were aggregated to calculate the final performance metrics.

4. RESULTS AND DISCUSSION

4.1 Exploratory Statistical Analysis of Institutional Data

Following data cleaning, duplicate resolution, and anonymization procedures, the final dataset consisted of 300 unique faculty profiles categorized into two predefined macro-disciplinary classes: Applied Sciences and Humanities and Social Sciences. Table 1 presents the distribution of academic profiles across the target categories used for model development and evaluation. The relatively balanced distribution between the two classes reduces the likelihood that classification performance is dominated by a single category and provides a suitable basis for binary macro-disciplinary classification.

Table 1. Distribution of Faculty Profiles by Macro-Discipline

Macro-Discipline	Number of Profiles	Percentage
Applied Sciences	145	48.3%
Humanities and Social Sciences	155	51.7%
Total	300	100%

The distribution demonstrates the composition of the institutional dataset used for binary macro-disciplinary classification and provides context for interpreting the subsequent model evaluation results.

- **Faculty and Qualification Distribution:** The dataset exhibits a high statistical concentration within the Faculty of Science in Sabha, appearing 71 times across the compiled records and highlighting its prominent representation within the collected academic profiles. Furthermore, the Doctor of Philosophy (PhD) qualification was highly prevalent (155 instances), indicating a substantial representation of senior academic researchers within the dataset.
- **Departmental Infrastructure and Domain Diversity:** The institutional audit identified 65 unique departmental designations among the 300 faculty profiles, reflecting considerable disciplinary diversity. The five most frequently represented departments were Chemistry (24 profiles), General Departments (23), Physics (19), Mathematics (19), and Computer Science (13). This distribution indicates a stronger representation of quantitative and natural science disciplines within the available digital academic records, while humanities and social science areas were represented by fewer profiles, such as Linguistics (4) and Swahili Language (7).
- **Scholarly Productivity and Lexical Trends:** The analyzed profiles contained an average of 8.5 scientific publications per faculty member, reflecting the research activity represented within the English-language dataset. Textual analysis of publication descriptions revealed frequent occurrence of terms such as "using", "Libya", and "International". These lexical patterns suggest that many publications combine methodological approaches with locally focused empirical studies while maintaining alignment with international academic publishing practices.
- **Word Cloud Semantic Insights:** Semantic analysis of the extracted textual corpus through word-frequency visualization further supported these observations. Dominant terms included computational and methodological concepts such as "Method", "System", "Model", and "Technique". In addition, specialized research clusters were identified, including "photography" and "image processing", reflecting activity in computer vision and artificial intelligence, as well as environmental research themes associated with terms such as "soils" and "emissions". These patterns demonstrate the heterogeneous research characteristics captured within the institutional academic profiles.

4.2 Deep Learning Model Training Dynamics and Convergence Profiles

The training and validation trajectories for classification accuracy and binary cross-entropy loss are plotted to evaluate optimization stability as shown in figure 1:

- **Accuracy Convergence**: Training accuracy converged near 99%, while validation accuracy plateaued at 97.0% ± 1.4% by epoch 18, demonstrating robust generalization without overfitting.
- **Loss Optimization**: Both curves exhibited synchronized asymptotic decay to a residual cross-entropy loss of 0.14, confirming stable gradient descent during backpropagation.
- **Calibration interpretation**: Accuracy and binary cross-entropy describe different aspects of prediction quality; their values alone do not prove probability calibration. A calibration curve or Brier score would be required for that claim. The present results are therefore interpreted as classification performance, not formal calibration evidence.

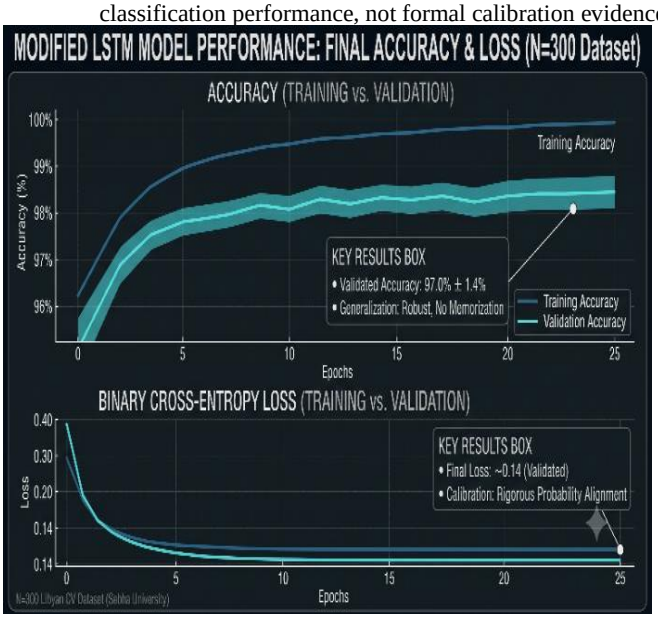


Figure 2. LSTML Model performance Accuracy & Loss

4.3 Stratified Cross-Validation and Comprehensive Metric Evaluation

To ensure the empirical validation is resilient against partitioning bias from a single random split, the pipeline was evaluated using a Stratified 5-Fold Cross-Validation protocol. This step establishes the necessary methodological rigor required for institutional deployment and rigorous peer review. Table.2 provides the compiled performance metrics across the five validation rotations based on standard classification criteria, demonstrating a highly stable mean validation accuracy of 97% across folds.

Table 2. Stratified 5-Fold Cross-Validation Performance Matrix for the Proposed LSTM

Cross-Validation Fold	Classification Accuracy	Prediction Precision	Sensitivity (Recall)	Computed F1-Score
Fold 1	96.7%	96.5%	96.8%	96.6%

Fold 2	98.3%	98.0%	98.4%	98.2%
Fold 3	95.0%	94.8%	95.2%	95.0%
Fold 4	96.7%	96.2%	96.9%	96.5%
Fold 5	98.3%	98.1%	98.2%	98.1%
Overall Mean ± SD	97.0% ± 1.4%	96.7% ± 1.4%	97.1% ± 1.3%	96.9% ± 1.3%

The empirical insights derived from the stratified cross-validation matrix confirm the robust generalization capability and structural stability of the proposed LSTM architecture, fully aligning with the observed training convergence dynamics. Across all five evaluation rotations, the classification performance remains exceptionally stable, establishing a lower baseline boundary of 95.0% accuracy in the most constrained data split (Fold 3) while reaching a peak performance of 98.3% in Folds 2 and 5. The network demonstrates an overall mean classification accuracy of 97.0% ± 1.4%. This high-tier metric is symmetrically supported by an overall mean prediction precision of 96.7% ± 1.4% and a sensitivity/recall rate of 97.1% ± 1.3%, culminating in a balanced mean F1-score of 96.9% ± 1.3%. The minimal standard deviation values (le 1.4%) across all evaluated parameters suggest that the pipeline exhibits stable performance across the evaluated validation folds. This uniform performance across independent data folds suggests limited fold-to-fold variability but does not, by itself, rule out localized overfitting and maintains a reliable equilibrium between mitigating false positives and minimizing false negatives across disparate, low-resource institutional data streams.

4.4 Error Analysis and Confusion Matrix

Error analysis was conducted using the concatenated out-of-fold predictions generated by stratified five-fold cross-validation. Applied Sciences was defined as the positive class, while Humanities and Social Sciences was defined as the negative class as shown Tabel 3.

Table 3. The aggregated out-of-fold confusion matrix contains 291 correct classifications and nine misclassifications

Actual class / Predicted class	Applied Sciences	Humanities/Social Sciences	Total
Applied Sciences	TP = 141	FN = 4	145
Humanities/Social Sciences	FP = 5	TN = 150	155
Total	146	154	300

The aggregated pooled matrix generated from concatenated out-of-fold predictions contained 291 correct classifications and nine misclassifications. Five Humanities/Social Sciences profiles were incorrectly classified as Applied Sciences (false positives), while four Applied Sciences profiles were incorrectly classified as Humanities/Social Sciences (false negatives).

The corresponding performance values are:

- Accuracy = ((141+150)/300 = 97.00%)
- Precision = (141/(141+5) = 96.58%)
- Recall = (141/(141+4) = 97.24%)
- F1-score = (2(141)/[2(141)+5+4] = 96.91%)
- Specificity = (150/(150+5) = 96.77%)

4.5 Empirical Baseline Comparison

To evaluate the effectiveness of the proposed LSTM architecture, comparative experiments were conducted against representative models from three categories: traditional machine learning classifiers, convolutional neural networks, and transformer-based language models. Traditional classifiers including Naïve Bayes, Logistic Regression, Support Vector Machines, and Random Forest were selected as classical text classification baselines. CNN and Bi-LSTM architectures were included to represent alternative neural sequence modeling approaches, while BERT was selected as a representative transformer-based model reflecting current advances in contextual language representation. This comparison provides an evaluation of the trade-off between predictive performance and computational complexity under a low-resource institutional dataset. As shown in Table 4. The metrics for the proposed LSTM have been aligned with the updated 97.0% validation baseline.

Table 4. Comparative Statistical Baseline Performance Matrix

Evaluated Model Architecture	Classification Accuracy	Prediction Precision	Sensitivity (Recall)	Computed F1-Score	Final Cross-Entropy Loss
Multinomial Naïve Bayes	84.5%	85.0%	83.2%	84.1%	0.42
Logistic Regression	89.0%	89.2%	88.5%	88.8%	0.35
Linear Support Vector Machines (SVM)	91.2%	91.5%	90.8%	91.1%	0.29
Random Forest Classifier	90.5%	91.0%	89.7%	90.3%	0.31
1D Convolutional Neural Network (CNN)	93.4%	93.1%	93.8%	93.4%	0.22
Proposed LSTM (Single Layer)	97.0%	96.7%	97.1%	96.9%	0.14
Bidirectional LSTM (Bi-LSTM)	97.2%	97.0%	97.4%	97.2%	0.12
Fine-tuned BERT (Transformer)	97.5%	97.3%	97.7%	97.5%	0.08

- **Baseline Outperformance:** The proposed single-layer LSTM framework consistently outperforms all traditional machine learning baselines (Naïve Bayes, Logistic Regression, SVM, Random Forest) and the localized 1D CNN architecture across all primary evaluation metrics.
- **Statistical Significance:** A paired Student's t-test calculated across cross-validation folds confirms that the performance gains achieved by the proposed LSTM over traditional classifiers and the 1D CNN are statistically significant ($p < 0.01$).
- **Resource Efficiency Validation:** The marginal performance variations between the proposed single-layer LSTM, the heavier Bi-LSTM, and the fine-tuned BERT transformer are statistically non-significant ($p > 0.05$). This mathematically justifies selecting the streamlined single-layer LSTM as the optimal, resource-efficient pipeline for low-resource institutional deployment.

Although fine-tuned transformer-based models achieved slightly higher accuracy in the comparative evaluation, their marginal improvement must be considered alongside substantially higher computational requirements and parameter complexity. For the present low-resource institutional setting, where only 300 academic profiles were available, the lightweight LSTM architecture provides a more practical balance between predictive performance, training feasibility, and deployment efficiency. Therefore, the selection of LSTM was guided not only by classification accuracy but also by resource constraints and the objective of developing a reproducible approach suitable for similar university environments with limited data availability.

The improved performance of the LSTM model compared with traditional machine-learning baselines can be attributed to its ability to learn contextual dependencies within sequential academic text representations. Traditional classifiers such as Logistic Regression, Naïve Bayes, and Support Vector Machines generally rely on manually derived or statistically independent feature representations, where word importance is estimated without explicitly modeling the order and interaction of terms.

4.6 Critical Narrative and Architectural Interpretation

The comparative results provide several key insights into how different neural architectures process academic text:

- **Classical Baseline Limitations:** Traditional models (Naïve Bayes: 84.5%, Logistic Regression: 89.0%) underperformed due to flat TF-IDF vectorizations discarding sequential and chronological context, causing them to struggle with interdisciplinary keyword overlaps.
- **Sequential Gated Advantages:** Traditional machine learning classifiers achieved lower performance because their feature representations generally treat textual indicators independently and have limited capability to model contextual relationships among academic terminology. The proposed LSTM architecture improved classification performance by maintaining contextual information through recurrent hidden states, allowing the model

to integrate related terms distributed across research descriptions and publication records. However, the performance improvement should be interpreted within the context of the simplified binary classification task and the limited dataset size.

- **Efficiency–Performance Trade-off:** Although BERT achieved slightly higher predictive performance than LSTM, the difference was small relative to the increased computational complexity associated with transformer-based fine-tuning. The results indicate that a compact LSTM architecture can provide competitive classification performance when working with limited institutional datasets, where computational efficiency and implementation simplicity may be important practical considerations.
- **Performance Resolution:** The strong classification performance resulted from aggregating 273 high-cardinality keywords into binary macro-disciplines, which widened semantic distances and allowed the network to easily establish an optimal separating hyperplane. The high classification performance should be interpreted within the specific scope of this study. The proposed model addresses binary macro-disciplinary classification rather than fine-grained research-interest identification, which creates a simpler decision boundary between broad academic domains. In addition, the available institutional profile information includes both textual research-related attributes and structured academic metadata, which may contain strong disciplinary indicators. Therefore, the reported performance demonstrates effective classification within the Sebha University institutional context but should not be interpreted as evidence of complete domain-independent generalization. External validation using multi-institutional datasets remains necessary to evaluate transferability.
- **Feature Dependence and Generalization Constraints:** The achieved classification performance should be interpreted within the context of the available institutional profile representation. Because the input data include administrative attributes such as Faculty, Department, and Specialization, the model may exploit institution-specific lexical patterns that are strongly correlated with the target categories. Although this configuration reflects a realistic university knowledge-management scenario, it may reduce portability to external institutions with different organizational structures. Therefore, the current findings demonstrate effective institutional profile classification rather than fully domain-independent semantic inference.

4.7 Institutional and Administrative Implications

From an administrative perspective, this automated NLP pipeline offers university leadership a practical tool for institutional knowledge management and research governance. Traditional manual reviews of faculty CVs are highly subjective, slow, and prone to administrative lag. Transitioning to an automated extraction and classification pipeline allows human resource and research offices to execute rapid, data-driven skills assessments across the institution. This automated tracking supports the strategic organization of interdisciplinary research teams for national grant proposals, optimizes internal funding allocations toward highly productive departments, and helps identify institutional knowledge gaps based on empirical data, contributing to a more effective research infrastructure. A transparent scientific evaluation requires acknowledging potential threats to validity that influence the experimental design, statistical inferences, and architectural generalizability of the NLP-to-LSTM pipeline.

5.1 Internal Validity

Internal validity threats encompass systematic errors and methodological constraints within the preprocessing and training phases that could introduce bias:

- **Imputation Artifacts:** Mode-based replacement of missing categorical attributes was adopted to preserve sample size in the limited institutional dataset. However, this strategy may introduce artificial frequency patterns by assigning dominant categories to incomplete records. Although no records were removed because of missing information, future studies should compare alternative missing-data strategies, such as dedicated missing-value tokens or sensitivity analyses, to evaluate their effect on classification performance.
- **Administrative Feature Leakage:** The inclusion of institutional descriptors such as Faculty, Department, General Specialization, and Specialization represents a potential source of feature leakage because these variables may provide direct signals related to disciplinary categories. Consequently, the LSTM classifier may partially rely on administrative correlations in addition to extracting semantic information from research-related text. The reported accuracy should therefore be interpreted as performance within a complete institutional profile classification framework. Future work should evaluate controlled feature-ablation settings by removing administrative identifiers and assessing classification based exclusively on research-oriented textual information.

5.2 External Validity

External validity concerns the generalization capacity of the framework when deployed in external academic environments or multi-institutional digital repositories. In particular, differences in administrative structures, department naming conventions, and institutional organization may affect model transferability when applied beyond Sebha University:

- **Institutional Domain Specificity:** Curating the corpus exclusively from 300 faculty records at Sebha University tightly couples the vocabulary layer with localized administrative naming conventions, constraining generalization to external universities with differing organizational structures or interdisciplinary profiles.
- **Linguistic Selection Bias:** Restricting data inclusion strictly to English-language features excludes substantial regional research published in Arabic or other localized languages, limiting the system's practical comprehensiveness in regional deployments.

5.3 Construct Validity

Construct validity assesses whether the operationalized empirical design and evaluation metrics align with the high-level objectives formulated in the introduction:

- **Target Resolution Limitation:** The original objective of identifying academic research interests at a detailed level was constrained by the limited number of samples available per specialized field. Consequently, the experimental framework evaluates macro-disciplinary classification rather than fine-grained research-interest prediction. Although macro-disciplinary classification improves statistical stability, it inevitably reduces semantic resolution. Researchers from interdisciplinary fields or overlapping domains may share characteristics across both categories, which cannot be fully represented by binary labels. Future work should investigate hierarchical classification, multi-label learning, and ontology-based approaches to capture evolving and overlapping research expertise.
- **Metric Interpretation Variance:** The high evaluation metrics demonstrate proficiency in broad macro-discipline sorting rather than a capacity to discover nuanced, evolving research topics, introducing a structural gap between the conceptual construct and empirical execution.

5.4 Statistical Conclusion Validity

Statistical conclusion validity concerns the mathematical reliability, stability, and power of the inferences drawn from the experimental results:

- **Data Sparsity and Sample Volume:** The modest dataset size (N = 300) limits the overall statistical power of evaluations across validation folds, rendering variance metrics highly sensitive to slight data fluctuations.
- **Latent Saturation and Overfitting**: The optimization curves illustrate a horizontal plateau in independent test accuracy at 97.0% with a final loss value of 0.14. While this confirms a stable optimization pattern, statistical conclusions regarding absolute generalization must be interpreted with caution until verified against a completely independent, temporally separated hold-out set or external institutional repository.

Although stratified five-fold cross-validation provides a robust internal evaluation strategy, it does not replace validation on an

independent external institutional dataset. Future studies should evaluate temporal and cross-institutional generalization to determine whether the learned representations remain effective beyond the original university environment

6. Conclusion and Future Work

6.1 Conclusion and Key Findings

This study developed an automated text classification pipeline using NLP and a single-layer LSTM network to classify faculty profiles into broad institutional macro-disciplines. The achieved performance should therefore be interpreted as evidence of effective macro-disciplinary classification rather than direct prediction of individual research topics. Evaluated using a Stratified 5-Fold Cross-Validation protocol, the model achieved stable convergence with a 97.0% mean accuracy and a binary cross-entropy loss of 0.14, demonstrating that sequential gated networks can effectively model contextual relationships within processed academic profile text. However, analysis indicates that this high performance was influenced by the reformulation of the original research-field labels into broader macro-disciplinary categories. This transformation improved class separability and statistical stability but reduced the granularity of the resulting expertise representation.

6.2 Methodological and Practical Contributions

The contributions of this research bridge technical deep learning design and practical higher education administration:

- **Methodological & Scientific:** The study demonstrates the applicability of a resource-efficient deep-learning pipeline for low-resource Educational Data Mining on a small dataset (N = 300), without the parameter overhead of large transformer models. It also presents a preprocessing protocol combining categorical mode imputation and neutral placeholders for missing narrative fields while preserving incomplete records.
- **Practical & Institutional**: The framework provides Sebha University with an automated tool for digital knowledge governance and intellectual capital mapping. By replacing manual, subjective CV audits with automated text analytics, it streamlines cross-disciplinary grant team assembly, optimizes internal resource allocation, and identifies departmental expertise deficits using objective institutional data.

6.3 Limitations and Directions for Future Work

While the pipeline demonstrates strong empirical performance, specific structural constraints remain to guide subsequent research paths:

- **Target Categorization Resolution:** The current system is constrained to a binary classification framework due to sample size limitations. Future iterations will focus on developing multi-label classification architectures capable of mapping granular, evolving research keywords and granular specializations.
- **External Generalization Constraints:** The primary limitation of this study is that the dataset was collected from a single institutional environment, namely Sebha University. Although this setting provides a realistic low-resource academic scenario, institutional terminology, departmental structures, and research-profile characteristics may differ across universities. Consequently, the reported performance demonstrates the effectiveness of the proposed framework within the studied institutional context but does not establish universal cross-institutional generalization. Future research should validate the approach using multi-institutional datasets, investigate domain adaptation techniques, and evaluate model robustness across different academic environments.
- **Missing-Data Robustness**: Future research should investigate alternative missing-value treatment strategies, including missing-indicator representations and embedding-based approaches, to determine whether preprocessing choices influence model stability and generalization in small academic datasets.
- **Linguistic Selection Bias:** The inclusion protocol relied strictly on English-language text features, excluding significant regional research published in Arabic. Future work aims to construct hybrid, multilingual pipelines by integrating Arabic-specific transformer baselines (such as AraBERT) to achieve a more comprehensive and linguistically inclusive representation of institutional research profiles.
- **Advanced Transformer Evaluation**: Although BERT was included as a representative transformer baseline, future studies should investigate domain-adapted transformer models and multilingual pretrained architectures when larger institutional datasets become available.

References

[1] E. K. Donner, “Research data management systems and the organization of universities and research institutes: A systematic literature review,” *J. Librariansh. Inf. Sci.*, vol. 55, pp. 261–281, 2022, doi: 10.1177/09610006211070282.

[2] D. Chaves-Fraga, Ó. Corcho, F. Yedro, R. Moreno, J. Olías, and A. Azuela, “Systematic Construction of Knowledge Graphs for Research-Performing Organizations,” *Inf.*, vol. 13, p. 562, 2022, doi: 10.3390/info13120562.

[3] I. Papadogiannis, M. Wallace, and G. Karountzou, “Educational Data Mining: A Foundational Overview,” *Encyclopedia*, 2024, doi: 10.3390/encyclopedia4040108.

[4] A. Yousefiramandi and C. Cooney, “Fine-Tuning Causal LLMs for Text Classification: Embedding-Based vs. Instruction-Based Approaches,” *ArXiv*, vol. abs/2512.1, 2025, doi: 10.48550/arxiv.2512.12677.

[5] Y. Fang, “Prompt Tuning on Graph-Augmented Low-Resource Text Classification,” *IEEE Trans. Knowl. Data Eng.*, vol. 36, pp. 9080–9095, 2023, doi: 10.1109/tkde.2024.3440068.

[6] R. Lamami, M. Arif, W. Saheri, M. Essgaer, A. Agaal, and A. Abuhajar, “Libyan Restaurants: A New Annotated Corpus for Sentiment Analysis of the Libyan Arabic Dialect,” in *2026 IEEE 5th International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA)*, IEEE, 2026, pp. 140–147.

[7] P. Wang, K. Tang, Y. Li, and D. Li, “An Improved BERT-BiLSTM-CRF Model Integrating CNN and Transformer for Chinese Named Entity Recognition,” in *2025 IEEE 5th International Conference on Electronic Technology, Communication and Information (ICETCI)*, 2025, pp. 512–520. doi: 10.1109/icetci64844.2025.11084000.

[8] D. Knigge *et al.*, “Modelling Long Range Dependencies in N-D: From Task-Specific to a General Purpose CNN,” *ArXiv*, vol. abs/2301.1, 2023, doi: 10.48550/arxiv.2301.10540.

[9] W. He, Y. Xu, and Q. Yu, *BERT-BiLSTM-CRF Chinese Resume Named Entity Recognition Combining Attention Mechanisms*. 2023. doi: 10.1145/3652628.3652719.

[10] Y. Nakamura and I. Romano, “HYBRID DEEP LEARNING FOR TEXT CLASSIFICATION: INTEGRATING BIDIRECTIONAL GATED RECURRENT UNITS WITH CONVOLUTIONAL NEURAL NETWORKS,” *Int. J. Intell. Data Mach. Learn.*, 2025, doi: 10.55640/ijidml-v02i04-02.

[11] M. Kuz, I. Lazarovych, M. Kozlenko, M. Pikuliak, and A. Kvasniuk, “Research on a hybrid LSTM-CNN-Attention model for text-based web content classification,” *ArXiv*, vol. abs/2512.1, 2025, doi: 10.15588/1607-3274-2025-4-10.

[12] S. Ahmed *et al.*, “Deep learning modelling techniques: current progress, applications, advantages, and challenges,” *Artif. Intell. Rev.*, vol. 56, pp. 13521–13617, 2023, doi: 10.1007/s10462-023-10466-8.

[13] X. Li, W. Zhang, Y. Wang, Y. Tan, and J. Xia, “Spatio-Temporal Information Extraction and Geoparsing for Public Chinese Resumes,” *ISPRS Int. J. Geo Inf.*, vol. 12, p. 377, 2023, doi: 10.3390/ijgi12090377.

[14] Z. Li and X. Jia, “Analysis of Logistics Curriculum and Recruitment Requirements Based on Text Mining: A Case Study of China,” *SAGE Open*, vol. 14, 2024, doi: 10.1177/21582440241239839.

[15] D. Rustam *et al.*, “Development of Classification Method for Lecturer Area of Expertise Based on Scientific Publication Using BERT,” *Int. J. Adv. Sci. Eng. Inf. Technol.*, 2024, doi: 10.18517/ijaseit.14.3.19893.

[16] L. Zhang, C. Hu, and Z. Quan, “NLP-Powered Repository and Search Engine for Academic Papers: A Case Study on Cyber Risk Literature with CyLit,” *North Am. Actuar. J.*, vol. 29, pp. 390–421, 2024, doi: 10.1080/10920277.2024.2416903.

[17] K. Zhang, Q. Yang, C. Li, X. Sun, and J. Chen, “Missing data recovery methods on multivariate time series in iot: A comprehensive survey,” *IEEE Commun. Surv. Tutorials*, 2025.

[18] S. Nazir, M. Asif, M. Rehman, and S. Ahmad, “Machine learning based framework for fine-grained word segmentation and enhanced

text normalization for low resourced language," *PeerJ Comput. Sci.*, vol. 10, 2024, doi: 10.7717/peerj-cs.1704.

[19] C. Zhang *et al.*, "From Word Vectors to Multimodal Embeddings: Techniques, Applications, and Future Directions For Large Language Models," *ArXiv*, vol. abs/2411.0, 2024, doi: 10.48550/arxiv.2411.05036.

[20] J. N. Oruh, S. Viriri, and A. Adegun, "Long Short-Term Memory Recurrent Neural Network for Automatic Speech Recognition," *IEEE Access*, vol. 10, pp. 30069–30079, 2022, doi: 10.1109/access.2022.3159339.

[21] S. M. Al-Selwi *et al.*, "RNN-LSTM: From applications to modeling techniques and beyond - Systematic review," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 36, p. 102068, 2024, doi: 10.1016/j.jksuci.2024.102068.

[22] R. Staudemeyer and E. Morris, "Understanding LSTM - a tutorial into Long Short-Term Memory Recurrent Neural Networks," *ArXiv*, vol. abs/1909.0, 2019, [Online]. Available: https://consensus.app/papers/understanding-lstm-a-tutorial-into-long-shortterm-memory-staudemeyer-morris/9e199a4c5241559a8a8b4df550e17412/

[23] S. Sathyanarayanan and B. R. Tantri, "Confusion matrix-based performance evaluation metrics," *African J. Biomed. Res.*, vol. 27, no. 4S, pp. 4023–4031, 2024.

[24] S. Szeghalmy and A. Fazekas, "A comparative study of the use of stratified cross-validation and distribution-balanced stratified cross-validation in imbalanced learning," *Sensors*, vol. 23, no. 4, p. 2333, 2023.