D-LexeCan: A Dynamic Lexicon-Based Framework for Sentiment Analysis in Tarifit, a Low-Resource Multiscript Language

Amar AMAKSSOUM, Fadwa BOUHAFER, Anass EL HADDADI, Abdelkhalak BAHRI
Data Science and Competitive Intelligence Team (DSCI) ENSAH, Abdelmalek Essaadi University, Tetouan, Morocco

Abstract*—Sentiment analysis for low-resource languages remains challenging due to limited annotated data, orthographic instability, informal writing practices, and the lack of dedicated linguistic resources, challenges that are particularly acute for Tarifit (Tamazight of the Rif), an under-resourced Amazigh language characterized by strong dialectal variation, pervasive multi-script usage, and highly noisy user-generated content on social media. This study introduces D-LexeCan, a dynamic lexicon-based sentiment analysis framework that infers polarity directly from annotated corpus evidence without relying on predefined sentiment dictionaries or computationally intensive pretrained deep learning and transformer-based models. The framework combines deterministic multi-script normalization, unifying Arabic script, Tifinagh, and Arabizi into a single Tarifit Latin representation with automatic induction of sentiment-bearing unigrams and bigrams, while explicitly modeling negation and amplification phenomena through linguistically motivated operators and preserving emojis as meaningful discourse-level sentiment cues. The approach is evaluated on a manually annotated social media corpus collected from multiple online platforms, where it achieves an accuracy of 0.8800 and a Macro-F1 score of 0.8798. The results outperform a static lexicon baseline with an accuracy of 0.5275, a classical machine-learning model based on TF–IDF and SVM with an accuracy of 0.8525, and neural architectures including BiLSTM with an accuracy of 0.7950. Experiments with frozen multilingual transformer encoders show accuracy ranging from 0.6725 to 0.7650. Fine-tuned multilingual transformers such as mBERT achieve competitive performance, reaching an accuracy of 0.8175. Overall, the results demonstrate that adaptive and linguistically grounded dynamic lexicon induction constitutes an effective, interpretable, and computationally efficient alternative for sentiment analysis in low-resource, noisy, and multi-script African language contexts.

***Keywords*—Sentiment analysis; low-resource languages; dynamic lexicon-based; static lexicon baseline; deep learning; machine-learning; transformer-based models**

I. INTRODUCTION

Sentiment analysis, or opinion mining, is a basic task in Natural Language Processing (NLP) that can be used in many ways, such as to keep track of public opinion [1], analyze social media [2], [3], and understand user feedback. In the last ten years, the field has changed from static lexicon-based and rule-driven methods [4], [5], to machine learning and deep neural architectures that are data-centric. This is mostly because large annotated corpora and pretrained transformer models that can capture contextual and semantic representations are now available. Because of this, architectures like BERT, T5, RoBERTa, and their multilingual versions have set the bar for performance on many sentiment analysis benchmarks [6], [7]. But these improvements have mostly happened in languages with a lot of resources, which means that many African and minority languages are still at a big disadvantage when it comes to performance, accessibility, and robustness.

The current situation reveals itself most clearly through sentiment analysis of low-resource languages because these languages face significant challenges when developing models, as they lack proper annotation datasets and necessary linguistic tools and standardized writing systems [8], [9]. The current multilingual pretrained language models XLM-R and AfroXLM-R, AfriBERTa and AraBERT, DarijaBERT, and mDeBERTaV3 demonstrate that transfer learning methods reduce the need for large datasets during testing on the shared AfriSenti benchmark. Still, these methods rely a lot on pretrained infrastructures and labeled data, often need a lot of computing power, and usually don't offer much insight. The system fails to effectively represent common digital communication patterns, which include multi-script variation and code-switching, expressive elongation, dialect-specific vocabulary, and non-standard tokens such as emojis and creative spellings [10], [11], [12].

These problems are even worse in Tarifit, a very under-resourced Amazigh language spoken in northern Morocco. Tarifit content created by users shows a lot of spelling mistakes in different writing systems, such as Latin script, Arabic script, Arabizi numerals, and Tifinagh. The text presents multiple dialects that it uses to communicate through casual language throughout its entire content. Sentiment is often expressed through elongated forms, discourse markers, amplifiers, negation constructions, and emojis, which together make the vocabulary less rich and make traditional NLP pipelines less effective. Without specialized sentiment lexicons, strong preprocessing tools, or pretrained models for specific languages, current sentiment analysis methods have trouble working well on real-world Tarifit data.

To mitigate these constraints, this research presents D-LexeCan, a dynamic lexicon-based sentiment analysis framework meticulously crafted to align with the linguistic and orthographic attributes of Tarifit. D-LexeCan is different from traditional lexicon-based methods that use predefined sentiment inventories. Instead, it starts with an empty lexicon and gradually adds sentiment polarity directly from user-generated content that has been annotated, using statistically sound evidence. The framework combines deterministic multi-script normalization into a single Tarifit Latin representation, automatic generation of sentiment-aware unigrams and bigrams to capture both lexical and multiword sentiment expressions, and operator-aware scoring systems that explicitly model negation and amplification effects, while keeping emojis as discourse-level sentiment cues.

We test D-LexeCan on a set of 2,000 user-generated comments from YouTube, Facebook, Instagram, and TikTok that have been manually annotated. The comments are divided into two groups: 80% for training and 20% for testing. The framework is evaluated against a static lexicon baseline, a traditional machine-learning model (TF–IDF + SVM), a deep learning model (BiLSTM), and various transformer-based configurations, including frozen multilingual embeddings and end-to-end fine-tuned models derived from mBERT and XLM-RoBERTa. Experimental results show that D-LexeCan strikes a good balance between predictive performance, interpretability, and computational efficiency. It beats all the baselines that were tested and does not have the high computational cost and lack of transparency that transformer-centric solutions do.

In general, this study shows that sentiment analysis can be done on low-resource, multi-script languages. Linguistically grounded lexicon-based modeling can help African languages a lot more than just using static lexical resources or computationally expensive pretrained transformers. The D-LexeCan system provides a transferable methodological framework that helps them solve three main problems that affect under-resourced languages, which include dealing with multiple writing systems, changing digital communication patterns, and limited access to NLP resources. The system operates beyond Tarifit language boundaries.

II. Related Work

Research on sentiment analysis has evolved through different analytical approaches, which started with lexicon-based methods and then moved to machine learning, deep learning, and finally transformer-based models. The methods have achieved major progress for high-resource languages, yet they fail to deliver effective results when working with low-resource and African languages because these languages lack sufficient data and present diverse writing systems and multiple languages. The following section establishes the research context through a review of previous studies that studied sentiment analysis under restricted data availability conditions.

The Dynamic lexicon-based sentiment analysis method functions as an alternative to standard fixed sentiment dictionaries, which track language polarity shifts between different contexts in newly collected text data. The research field requires Mechulam et al. [13] created a dynamic lexicon system which uses graph-based structures and valence propagation methods to detect sentiment polarity through text usage analysis. The method achieves successful results for domain adaptation and context-sensitive polarity modeling, but it needs extensive annotated data and complicated propagation rules, and it depends on spelling consistency. The system operates at its best when it encounters linguistic environments that contain plenty of resources and use standardized language patterns. Saileela and Malleswara Rao [14] created a dynamic lexicon-based sentiment analysis system, which applied nonlinear optimization and genetic algorithms to enhance n-gram and sentiment phrase feature selection. Their framework improves sentiment prediction accuracy through the combination of evolutionary optimization with phrase concurrence analysis and machine-learning classifiers. The system depends on pre-established sentiment dictionaries, yet it needs extensive data preparation and substantial processing resources to analyze structured monolingual information, which does not include unstructured multilingual user content.

Aliyu et al. [8] conducted a systematic literature review which applied SLR/PRISMA protocol to study sentiment analysis in low-resource environments and delivered an extensive assessment of this topic. The review narrows down 225 candidate studies to 56 publications from 2018 to 2023 and looks at them in terms of research questions, methods, data sources, preprocessing methods, feature extraction methods, and evaluation metrics. The three primary paradigms in this field consist of machine translation, word embeddings, and transfer learning with pretrained transformer models. The research depends on Twitter and Facebook data for its analysis, while accuracy and F1-score serve as the primary evaluation metrics. The team organized standard preprocessing operations together with feature encoding methods and current challenges, which include insufficient language processing capabilities and code-switching, sparse data distribution, and unequal resource availability. The review is mostly descriptive, doesn't put together a real end-to-end system, and mostly talks about general deep-learning models instead of lexicon-centric or dynamic lexicon architectures made for a single language that is very under-resourced and has a lot of spelling and dialectal differences.

Research has concentrated on applying Transformer-based models for analyzing sentiment across multiple languages, which include African languages. The research by Raychawdhary et al. [10] evaluated large-scale pretrained architectures like XLM-R, AfroXLM-R, AfriBERTa, and mDeBERTaV3 using AfriSenti SemEval datasets that include Hausa, Yoruba, Darija, and Swahili. The proposed framework consists of five components, which include data selection and normalization, SentencePiece tokenization, and multilingual fine-tuning and zero-shot transfer learning to classify tweets into positive, negative, and neutral sentiment categories. The results show that transformer-based transfer learning works well for very low-resource languages, especially with AfroXLM-R. However, the method mostly uses pretrained multilingual representations and doesn't explicitly model linguistic phenomena that are relevant to sentiment, like negation, emojis, dialectal noise, or multiscript variation. It also still relies on manually labeled corpora instead of dynamically induced lexical knowledge. Koto et al. [15] presented a related zero-shot strategy by pretraining multilingual language models on an extensive NRC-VAD lexicon filtered across over one hundred languages. The method shows improvements over standard fine-tuning and

English-only transfer by learning valence-based sentiment labels from lexicon entries and using the models directly on sentence-level tasks in 34 languages, including some African languages. However, it still relies on large, manually created high-resource lexicons, uses mostly English-based extension strategies, and works with static lexical resources instead of sentiment lexicons that change over time based on real user-generated content.

Researchers have also looked into supervised transformer-based sentiment analysis for single low-resource languages. For Central Kurdish, Awlla et al. [16] looked at a BERT-based framework that was trained on a normalized corpus with a dedicated tokenizer. They also tested different architectures, such as BiLSTM, MLP, and end-to-end BERT fine-tuning. Their findings indicate that contextualized embeddings significantly surpass conventional Word2Vec representations, with a fine-tuned BERT attaining an accuracy of 75.37% in three-class classification and 86.31% in the binary context. Even with these improvements, the framework is still completely supervised and model-centric. It uses pretrained contextual embeddings and annotated data, and it doesn't explicitly model sentiment-specific linguistic phenomena like negation, amplification, emoji semantics, or multiscript variation that are common in informal user-generated content.

Several studies have examined cross-lingual deep-learning frameworks for Uyghur sentiment analysis. One study integrates LaBSE sentence embeddings with a BiLSTM layer and utilizes AEDA-based data augmentation to address the issue of annotation scarcity. The system does better than a number of multilingual baselines on binary and multi-class datasets, but it is still fully supervised and relies on external pretrained models and augmented labeled data. It does not explicitly model or handle negation, intensification, emojis, or orthographic variation factors that make it harder to understand and adapt in noisy user-generated contexts [17].

There have also been suggestions for deep-learning-based sentiment analysis pipelines that use word embeddings and neural classifiers for other languages with few resources. Ali et al. [18] investigated a framework for Urdu that involves the creation of large-scale datasets, comprehensive preprocessing, and feature learning utilizing Word2Vec, succeeded by CNN and RNN classifiers for multi-class sentiment prediction. The results show that deep neural models can work when there is enough well-organized data, but the method only works with one writing system and doesn't include lexicon induction, polarity propagation, or ways to deal with informal, changing, or multi-script user-generated language. Iqbal et al. [19] employed a similar supervised approach for Pashto social media sentiment analysis, integrating traditional text features with dense word embeddings and assessing a diverse array of conventional, ensemble, and deep-learning classifiers on a limited manually annotated Facebook corpus. Even though it works well in practice, the framework is still only for one language and one script. It doesn't have a sentiment lexicon or a way to grow the lexicon dynamically, and it doesn't explicitly model negation, amplification, or large-scale orthographic variability.

Lexicon-enhanced but static sentiment analysis methods have also been studied. Meetei et al. [20] created a multi-stage sentiment analysis framework for Manipuri that combines multi-script normalization, handcrafted negative-morpheme lexicons, traditional classifiers, and further experiments with CNN, LSTM, and BiLSTM models. The study emphasizes the significance of script normalization and language-specific sentiment markers; however, it concentrates on formal news text, depends extensively on manual engineering, and lacks support for dynamic sentiment adaptation as new expressions arise.

Lexicon-informed feature engineering has been used to look into semi-supervised sentiment classification. Liu et al. [21] proposed a framework that enhances conventional document-level classifiers by integrating features from a general-purpose sentiment lexicon, amalgamating TF-IDF vectors, sentiment vectors, and statistical aggregates. Lexicon-based features enhance classification performance; however, this methodology depends on a static sentiment resource, presumes the existence of high-quality linguistic tools, and fails to derive or adjust lexical polarities from noisy, code-mixed, or multi-script data. Dhananjaya et al. [22] looked into a lexicon-based intermediate-task fine-tuning method for multilingual pretrained language models that uses auxiliary phrase construction to add cross-lingual sentiment alignment signals. This is a step toward hybrid transformer-centric strategies. Even though the method consistently outperforms standard fine-tuning, it is still closely tied to transformer infrastructures and high-resource sentiment lexicons. It doesn't take advantage of dynamically evolving lexical knowledge from the target low-resource language itself.

Mohammed and Prasad [23] examined a particularly pertinent contribution to lexicon-based sentiment analysis for low-resource languages, presenting a thorough methodology for developing a sentiment model for Hausa. Their framework integrates lexicon compilation, regulated linguistic enhancement, expert annotation, and subsequent BERT fine-tuning, resulting in elevated accuracy on annotated Hausa tweets. The approach, however, depends on a static pre-built lexicon and doesn't deal with multiscript heterogeneity, spontaneous orthographic variation, or the dynamic emergence of new expressions. This means that more research is needed into adaptive lexicon-based frameworks for low-resource languages.

III. Methodology

This research presents D-LexeCan, a modular sentiment-analysis framework developed for Tarifit (Tamazight of the Rif), a low-resource language marked by scarce linguistic resources, significant dialectal diversity, and varied writing conventions. D-Lexicon is the most important part of the framework. It is a dynamic sentiment-induction component that learns polarity directly from annotated data instead of using pre-made sentiment dictionaries. D-Lexicon is different from traditional lexicon-based methods because it starts empty and then fills in and updates its entries based on how people actually use words in labeled comments. An overview of the complete D-LexeCan sentiment analysis pipeline is illustrated in Fig. 1. As Tarifit text is processed, each token, spelling variant, emoji, or multiword expression is statistically evaluated across positive and negative contexts, and polarity is assigned only when usage exhibits stable and meaningful asymmetry. The lexicon automatically changes polarity assignments when new forms, dialectal variants, or new online expressions come out. This is because evidence changes. This adaptive behavior is especially important in Tarifit, where people can use Latin script, Arabic script, Arabizi, and Tifinagh (ⵜⴰⵎⴰⵣⵉⵖⵜ) all at the same time in the same online spaces. D-LexeCan offers a clear and data-driven framework for sentiment analysis in situations where standardized linguistic resources are mostly missing. It does this by combining preprocessing, multiscript normalization, dynamic lexicon growth, scoring, hyperparameter tuning through grid search, and evaluation into one pipeline.

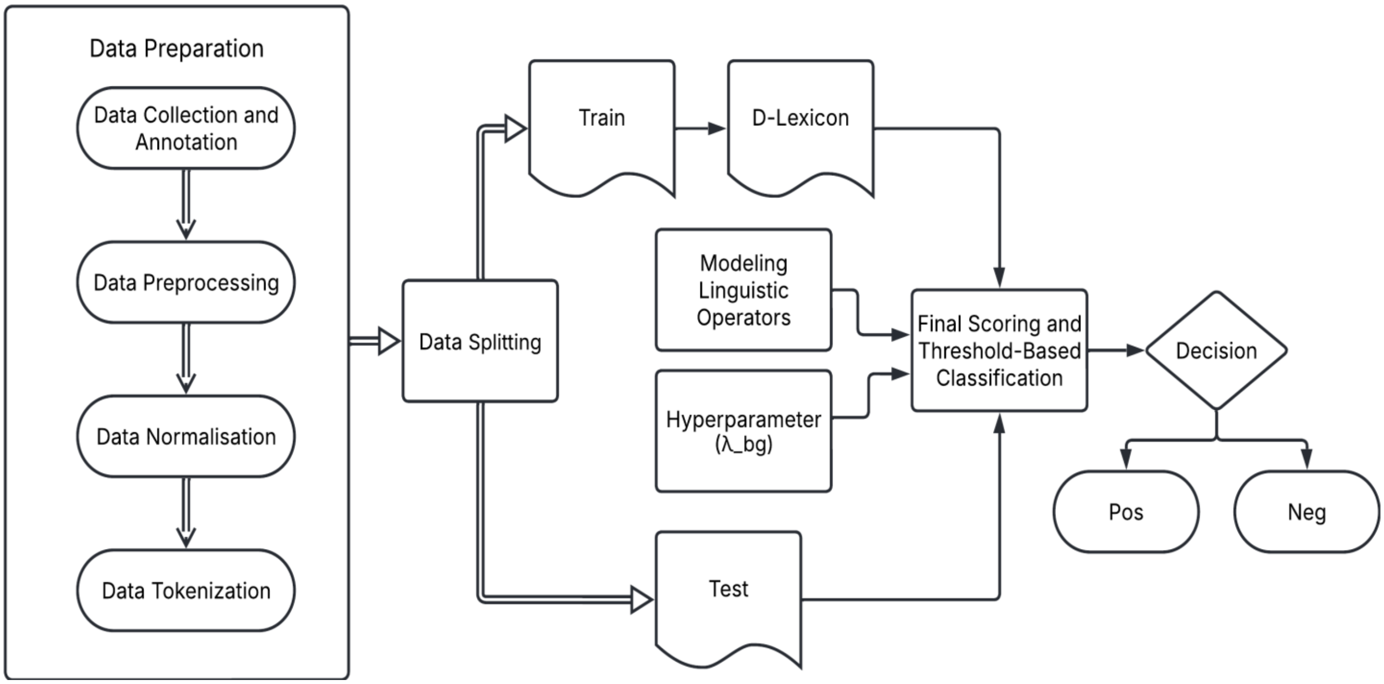

Fig. 1. Overview of the D-LexeCan sentiment analysis pipeline.

A. Data Preparation

The data preparation process transforms unprocessed user-generated content into an analysis-ready format through multiple operations, which start with data acquisition and human labeling, followed by orthographic noise reduction and writing system unification and word and emoji extraction through tokenization. The operations maintain consistent representation while keeping essential linguistic elements that affect sentiment, so the system can reliably create new words and analyze emotions.

1) Data collection and annotation from multi-platform, multi-script, and multi-dialect sources: The dataset consists of 2,000 user comments, including 1,000 positive and 1,000 negative statements. The comments were taken from popular online sites where Tarifit speakers are active, such as YouTube, Facebook, Instagram, and TikTok, and they covered a wide range of cultural, social, and entertainment topics. The corpus shows actual human communication patterns that people use when they talk to each other instead of using fabricated dialogue examples. The dataset is very multi-script, which is one of its most important features. For example, Tarifit users often use Arabizi, Arabic script, Latin-based Tarifit, and Tifinagh all in the same discussion thread or even in the same comment. The corpus also shows a lot of dialectal variation from different parts of the Rif, such as Al Hoceima, Nador, Driouch, Aknoul, Ayt Waryaghar, Ayt Touzin, Bni Bouayach, Temsamane, and the areas around them. This dataset is very representative of how Tarifit is actually used because it has many platforms, scripts, and dialects. However, it is also much harder to process. Native speakers of Tarifit manually annotated all of the comments, giving each one a binary sentiment label (Positive or Negative) based on its overall communicative intent and pragmatic interpretation. The research needed this method to identify intricate emotional expressions that standard language systems fail to show. As far as we know, the resulting corpus is one of the first balanced and carefully curated sentiment datasets for Tarifit.

2) Data preprocessing: The dataset goes through a normal preprocessing phase after it has been gathered and marked up. This phase tries to fix spelling mistakes and make sure that all sorts of user-generated material look the same. To keep words like "Zin," "ZIN," and "zin" from being repeated on the surface, all comments are first transformed to lowercase. The next step is to use a regulated procedure to get rid of dirt and other pollutants on the surface. In this process, only alphanumeric characters, whitespace, and emojis that show positive or negative feelings are kept. All other symbols, punctuation marks, and random Unicode artifacts are replaced with spaces. To further reduce spelling differences and expressive elongation that are common in social media conversations, sequences of repeated characters are normalized by combining any three or more identical characters into exactly two (e.g., atasss → atass, hhhhh → hh). Whitespace normalization finally gets rid of extra spaces and makes sure that token boundaries are clear. All of these preprocessing steps work together to reduce orthographic variability, vocabulary fragmentation, and create a stable text representation that can be used for multiscript normalization, tokenization, and sentiment modeling.

3) Multiscript normalization: To deal with the very different ways that people write Tarifit, the framework uses a full multiscript normalization process that turns all surface forms into a single Tarifit Latin (TL) representation. This step is necessary to combine script-induced variants into a single canonical space and stop artificial vocabulary inflation because comments are written in Tifinagh, Arabic script, Arabizi, and inconsistent Latin transcriptions.

Normalization is done in a set order so that transformations are always the same and can be repeated:

- Tifinagh→TL Tifinagh characters are transliterated into their TL phonemic equivalents using a deterministic mapping (e.g., ⵅ → x, ⵄ → ε, ⵓ → u, ⵥ → ẓ).
- Arabicscript → TL Arabic characters are converted through consistent one-to-one mappings reflecting their phonological role in Tarifit (e.g., ش → c, خ → x, غ → ɣ, ق → q, ط → ṭ).
- Arabizi→TL Digit–letter substitutions (e.g., 3 → ε, 7 → ḥ, 9 → q) and informal digraphs (e.g., ch/sh → c, gh → ɣ, kh → x) are normalized to standardized TL forms.

  Representative examples of multiscript normalization into Tarifit Latin are shown in TABLE I.

TABLE I. ILLUSTRATIVE EXAMPLES OF MULTI-SCRIPT NORMALIZATION INTO TARIFIT LATIN

Original Text	Normalized Tarifit Latin
ⵡⴰⵍⵍⴰⵀ ⵉⵍⴰ ⵖⴰⵛ ⵔⵃⴰⵇ	wallah ila ɣac rḥaq
نش بعدا ثعجباي	nc bɛḏa ṯɛğby
Bazz ichek awma	bazz icek awma

This stage directly helps downstream dynamic lexicon induction by enforcing a single TL representation. This is because dynamic lexicon induction depends on consistent token identities instead of script-dependent surface variants.

4) Tokenization: After character-level preprocessing and multiscript normalization, each comment is segmented into minimal linguistic units using a regular-expression-based tokenizer specifically adapted to the requirements of the D-LexeCan framework. The tokenizer operates on normalized Tarifit Latin (TL) text and extracts 1) word tokens as contiguous sequences of word characters (pattern \w+) and 2) standalone emojis treated as independent tokens during tokenization. Because tokenization is applied after both character cleaning and script harmonization, reusing the same normalization procedure, all non-emoji tokens strictly follow the unified TL representation and no longer contain residual symbols, script artifacts, or non-standard encodings. Emojis are preserved as standalone tokens rather than removed or treated as auxiliary information. The system stores the resulting token sequences in a Tokens_norm field, which exists for each comment to create a standardized format that can be used again. The D-Lexicon system defines tokens as an extensive category that encompasses standardized words and dialectical expressions and single emoji symbols that undergo the same automated processing as text elements.

5) Train–test partitioning: The evaluation process requires a rigorous and unbiased assessment, which uses the preprocessed and annotated corpus to create training and testing subsets through an 80/20 stratified split with SEED = 42 for reproducibility purposes while maintaining the 50/50 positive to negative label ratio in both subsets. The training data serves only to create the dynamic lexicon and sentiment-bearing bigram patterns and to calculate token and bigram frequency statistics, polarity confidence ratios, and model hyperparameter optimization, including min_freq, min_conf, and λ_bg through multiple testing of different options. The test subset stays hidden from view throughout both induction and calibration stages because it serves only for the final assessment of D-Lexicon and all comparison models, which prevents overfitting and memorization.

B. D-Lexicon: Dynamic Sentiment Induction Core

D-Lexicon constitutes the central adaptive component of the proposed D-LexeCan framework. The system functions as an entire data-driven, dynamic sentiment lexicon that learns sentiment polarity from user-generated content annotations instead of using pre-defined sentiment dictionaries or external lexical resources. Unlike static lexicons, D-Lexicon is initialized as an empty structure and incrementally constructs its sentiment knowledge by observing how lexical units, words, emojis, and multiword patterns are used across positive and negative contexts in the training data. The overall construction process of the dynamic D-Lexicon component is illustrated in Fig. 2. The design of D-Lexicon allows it to adapt continuously to the different orthographic forms and dialects and new words that appear in informal Tarifit speech while showing all sentiment evaluation processes.

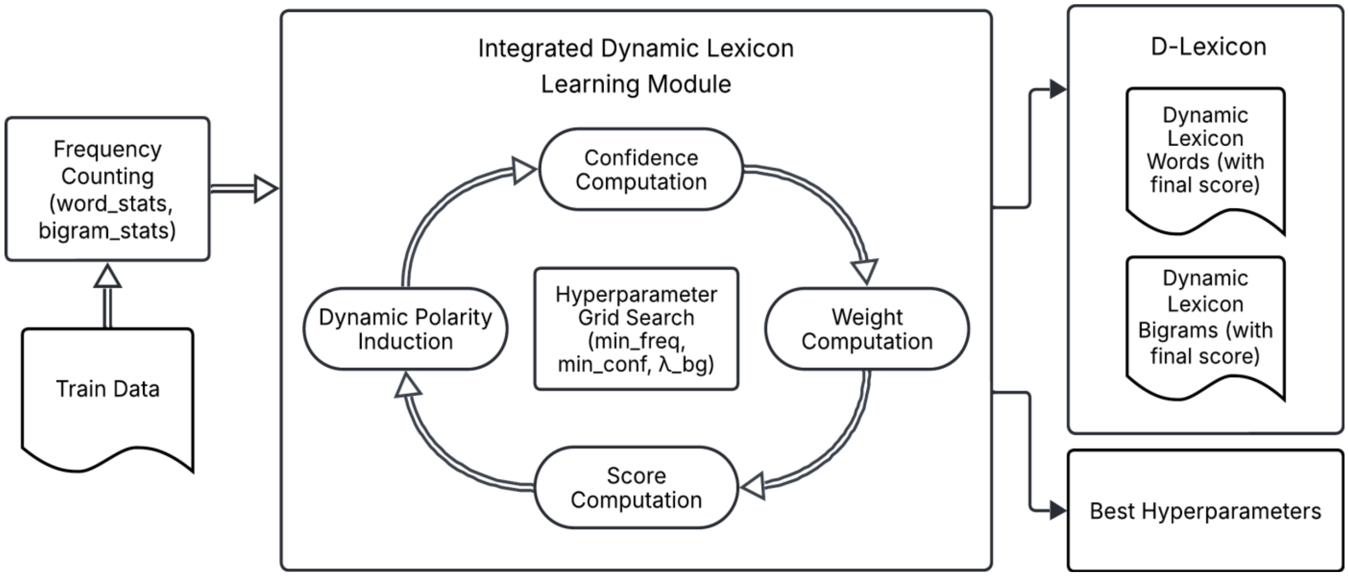

Fig. 2. Construction process of the D-Lexicon dynamic sentiment lexicon.

1) Hyperparameter configuration and optimization: The system D-Lexicon operates through a limited number of hyperparameters, which determine how to balance between word coverage and system stability and context-based performance during its dynamic lexicon creation process. The minimum frequency threshold (min_freq) in the system establishes the number of occurrences that tokens and sentiment-bearing bigrams need to meet for analysis, while it removes unimportant forms that occur rarely in informal Tarifit speech. The polarity confidence threshold (min_conf) establishes the minimum confidence level that enables tokens to obtain fixed sentiment orientation for entering the lexicon. The bigram scaling factor (λ_bg) serves as an additional control that determines how much bigrams with sentiment value should influence the model in relation to unigrams, so they can enhance word-level sentiment indicators without taking over their importance. The optimal hyperparameter configuration is identified through grid search on the training set and fixed before evaluation on the held-out test set.

2) Dynamic polarity induction, automatic lexicon construction, and lexicon rebuilding: For each candidate hyperparameter configuration (min_freq, min_conf) explored during grid search, D-Lexicon induces sentiment knowledge directly from the annotated training corpus, without relying on any predefined sentiment dictionary. The training data contains all observed tokens t, which include standard words and dialectal variants, non-standard spellings, and emojis for potential sentiment analysis.

a) Token frequency estimation: Each token t is associated with two class-conditional occurrence-level frequency counters, initialized as:

$$f_{pos}(t) = 0, f_{neg}(t) = 0 \quad (1)$$

where, f_pos (t) denotes the total number of occurrences of token in positive-labeled comments, while f_neg (t) the total number of occurrences of (t) appears in negative-labeled comments. As the labeled training data are processed, counters are incremented for each occurrence of token t according to the sentiment label of the comment in which it appears. The system design of repeated token occurrences in single comments maintains their multiple counts because it keeps essential information about user-generated content intensity and repetition, which appears in informal user-generated content.

The total frequency of token t across the corpus is then computed as:

$$f_{tot}(t) = f_{pos}(t) + f_{neg}(t) \quad (2)$$

Where f_tot (t) reflects the overall usage strength of token t in the training data.

To ensure robustness and reduce noise, two statistical constraints are enforced.

b) Minimum frequency constraint

$$f_{tot}(t) \geq min\_freq \quad (3)$$

This constraint removes spurious or accidental tokens whose total number of occurrences is too small to support reliable sentiment induction.

c) Polarity confidence constraint

$$Conf(t) = max\left(\frac{f_{pos}(t)}{f_{tot}(t)}, \frac{f_{neg}(t)}{f_{tot}(t)}\right), Conf(t) \geq min\ \_conf \quad (4)$$

This confidence value measures the degree of sentiment dominance exhibited by token t, quantifying how strongly its occurrences are biased toward one sentiment class. Tokens with low confidence correspond to sentiment-ambiguous forms, whereas high-confidence tokens exhibit stable sentiment orientation.

Only tokens satisfying both constraints are retained in the dynamic lexicon.

d) Polarity assignment: For each retained token, sentiment polarity is assigned according to the dominant class:

$$Polarity = \begin{cases} +1, & if\ f_{pos}(t) > f_{neg}(t) \\ -1 & otherwise \end{cases} \quad (5)$$

e) Frequency-aware weighting and token scoring: A confidence-weighted magnitude is computed for each retained token:

$$Weight(t) = Conf(t) \times ln\left(1 + f_{tot}(t)\right) \quad (6)$$

The final sentiment score stored in the lexicon is defined as:

$$Score(t) = Polarity(t) \times Weight(t) \quad (7)$$

This continuous score jointly encodes sentiment direction, statistical reliability, and occurrence-based usage strength in a single scalar value.

3) Multiword pattern learning: bigram-level sentiment induction: In addition to single-word entries, and under the same candidate hyperparameter configuration (min_freq, min_conf, λ_bg), D-Lexicon also induces sentiment information from contiguous word pairs (bigrams) whose polarity emerges only through co-occurrence. In informal and low-resource languages such as Tarifit, sentiment is frequently conveyed via such local multiword patterns, for instance, when an otherwise neutral word acquires polarity in a specific context or when sentiment intensity is reinforced through short lexical sequences.

During training, bigrams are extracted sequentially from each normalized comment. Occurrence-based counting is applied for bigrams, meaning that repeated occurrences of the same bigram within a single comment are counted multiple times. This choice preserves emphasis effects and repeated sentiment expressions typical of user-generated discourse.

For each bigram $bg = (ti, ti + 1)$, two class-conditional occurrence-level frequency counters are maintained: $f_{pos}(bg)$ which denotes the total number of occurrences of the bigram in positive-labeled comments, and $f_{neg}(bg)$ which represents the total number of occurrences in negative-labeled comments. The total frequency of the bigram is then defined as:

$$f_{tot}(bg) = f_{pos}(bg) + f_{neg}(bg) \quad (8)$$

Minimum frequency filtering, polarity confidence estimation, and polarity assignment follow the same definitions as for unigrams.

Unlike unigrams, the contribution of multiword patterns is explicitly controlled through a bigram scaling factor λ_bg, reflecting their higher semantic specificity and more contextual nature.

For each retained bigram, the confidence-weighted magnitude is computed as:

$$Weight(bg) = \lambda\_bg \times Conf(bg) \times ln(1 + f_{tot}(bg)) \quad (9)$$

The final sentiment score assigned to a bigram is then defined as:

$$Score(bg) = Polarity(bg) \times Weight(bg) \quad (10)$$

This formulation allows bigrams to complement word-level sentiment cues without overwhelming them, while still capturing local compositional sentiment phenomena that cannot be inferred from individual tokens alone.

In the implementation, this mechanism corresponds directly to the bigram_stats structure and the dynamic_patterns dictionary produced by the build_dynamic_lexicon function. The bigram induction process runs independently for each candidate hyperparameter configuration during grid search to maintain strict consistency, reproducibility, and statistical control.

When additional labeled training data becomes available, sentiment statistics are recomputed, and the dynamic lexicon is rebuilt accordingly.

C. Modeling Linguistic Operators

1) Amplifiers: Some expressions do not carry intrinsic sentiment polarity but instead modulate the intensity of nearby sentiment-bearing words. In Tarifit online discourse, this category includes strong amplifiers such as atass, as well as weaker forms such as dross and cwit. Within D-Lexicon, amplifiers are modeled through a local multiplicative mechanism applied during comment-level scoring: when an amplifier is detected immediately after a sentiment-bearing token, the score of that token is scaled accordingly. Let Score(t) denote the base sentiment score of the token t; the adjusted score is computed as:

$$Scoreamp(t) = \begin{cases} \alpha_{strong} \times Score(t), if\ a\ strong\ amplifier\ follows\ t \\ \alpha_{weak} \times Score(t), if\ a\ weak\ amplifier\ follows\ t \\ Score(t) \end{cases} \quad (11)$$

where $\alpha_{strong} > 1$ and $0 < \alpha_{weak} < 1$ are fixed scaling factors. In the implementation. This design reinforces or attenuates sentiment strength without introducing independent sentiment contributions.

2) Negation: Negation plays a distinct and non-additive role in sentiment composition, typically reversing polarity rather than merely weakening it. D-Lexicon models negation through explicitly observed lexical forms such as *wayeji, waṯeji,* as well as through the productive pattern *wa* + *token* + *ca*, which consistently emerges after normalization.

Two local inversion mechanisms are implemented:

- Lexical negation inversion: When a negation marker is encountered, the polarity of the immediately following sentiment-bearing token is inverted:

$$Score_{neg}(t) = -Score(t) \quad (12)$$

- Pattern-based negation inversion (wa … ca). For sequences of the form wa + m + ca, where m is a sentiment-bearing token, the polarity of m is flipped:

$$Score_{neg}(m) = -Score(m) \quad (13)$$

This operation is implemented by subtracting the dynamic lexicon score of the middle token whenever the pattern is detected.

Together, these mechanisms model negation as a local polarity inversion process, avoiding global penalties or handcrafted grammatical rules while remaining consistent with the data-driven scoring framework.

D. Final Scoring and Threshold-Based Classification

Once all lexical contributions and linguistic operators have been applied, D-Lexicon aggregates them into a single continuous sentiment score for each comment. Let S(C) denote the sentiment score of a comment C, computed as the sum of token scores (words and emojis) and learned bigram-pattern scores after local operator application:

$$S(C) = \sum_{t \in C} Score(t) + \sum_{bg \in C} Score(bg) \tag{14}$$

During aggregation, amplifiers and negation are applied locally at scoring time. Amplifiers scale the scores of neighboring sentiment-bearing tokens multiplicatively, while negation inverts the polarity of the affected token or pattern. These operators do not introduce independent additive terms but directly modulate existing lexical contributions. Because all scores derive from dynamically induced weights and data-driven operator rules, S(C) reflects not only sentiment polarity but also intensity and contextual modulation at the comment level.

In the present study, sentiment classification is formulated as a binary task (positive vs. negative), without an explicit neutral category. Accordingly, a fixed zero-threshold decision rule is adopted:

$$S(C) > 0 \Rightarrow Positive, S(C) \leq 0 \Rightarrow Negative \tag{15}$$

The case S(C) = 0 denotes the absence or exact cancellation of detectable polarity cues within a comment and is empirically observed to occur rarely in the evaluated dataset. As a strictly binary setting is assumed, all instances must be assigned to one of the two classes to ensure a complete decision function. Therefore, comments with S(C) = 0 are assigned to the negative class by convention, without implying the presence of explicit negative sentiment. The design choice prevents the automatic addition of a neutral class, which reduces the tendency to make optimistic predictions about positive results while keeping the system interpretable through its use of a single continuous sentiment score, which emerges from the dynamically created lexicon for each comment.

With the availability of larger annotated datasets, future work will explore the explicit incorporation of a neutral class corresponding to S(C) = 0 within a multi-class classification framework.

E. Evaluation Protocol and Assessment Metrics for D-LexeCan

D-LexeCan framework was tested through its performance assessment on data that they reserved for testing purposes after splitting their training data into 80% for training and 20% for testing, while maintaining equal numbers of positive and negative sentiment examples. The model learned all linguistic cues from the training data through its dynamic process of discovering new words, determining word polarity, bigram values, and scoring methods. The trained model generated continuous sentiment scores, which it applied to all test set comments that it had not seen before. The raw score data received binary sentiment labels through a specific zero-based decision threshold, which operated as a data-driven classification system without requiring human-defined decision rules. The evaluation protocol follows standard procedures for sentiment analysis of low-resource languages through lexicon-based methods, while they focus on achieving both interpretability and reproducibility. The research maintained methodological rigor through its application of proven evaluation metrics, which consisted of Accuracy and Precision, Recall and F1-score, and Macro-F1 and confusion matrix analysis. The assessment strategy, which uses multiple performance indicators, follows the recommendations from current low-resource sentiment analysis research, which supports the need to evaluate performance through multiple indicators for achieving both system robustness and fairness [24], [25].

1) Accuracy

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{16}$$

Where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively. Although accuracy may be misleading under class imbalance, it remains informative in this study due to the perfectly balanced dataset.

2) Precision

$$Precision = \frac{TP}{TP+FP} \tag{17}$$

Precision measures the reliability of positive predictions. High precision indicates a low rate of false positives, which is particularly important in low-resource settings characterized by noisy, dialectal, and orthographically inconsistent user-generated content.

3) Recall

$$Recall = \frac{TP}{TP+FN} \tag{18}$$

Recall quantifies the model's ability to correctly identify sentiment-bearing instances. It is especially relevant when capturing subtle or weak sentiment expressions is more critical than minimizing misclassification, as often observed in informal or code-mixed discourse.

4) F1-Score

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{19}$$

The F1-score provides a balanced measure by jointly accounting for false positives and false negatives, offering a robust indicator of overall classification quality in noisy sentiment analysis scenarios.

5) Macro-F1 Score

$$Macro - F1 = \frac{1}{N}\sum_{i=1}^{N}\frac{2\times Precision_i \times Recall_i}{Precision_i + Recall_i} \quad (20)$$

This metric assigns equal importance to each class and is widely recommended in low-resource sentiment analysis to mitigate class-specific bias. Even for balanced datasets, its reporting ensures consistency and comparability with prior benchmarks.

6) Confusion matrix: The confusion matrix provides a detailed view of misclassification patterns, revealing tendencies such as over-prediction of negative sentiment or failure to detect weaker polarity cues. This level of interpretability is particularly valuable in lexicon-based and low-resource frameworks, where understanding error behavior is as important as overall accuracy.

IV. Results and Discussions

A. Dynamic D-Lexicon Within D-LexeCan

D-LexeCan adds a dynamic lexicon induction component (D-Lexicon) to the static baseline to get around its inflexibility and low coverage. The system learns to identify sentiment patterns through this component, which starts with basic sentiment understanding and then advances to polarity detection from the annotated training data. We adjusted the hyperparameters to get the best performance while keeping rare linguistic expressions that are typical of minority languages. This is the best configuration:

μιν_φρεθ = 1, μιν_χονφ = 0.65, ⎿_βγ = 1.20

The system generated an induced lexicon that contained 9,325 entries that detected sentiment and included 3,974 unigrams and 5,351 bigrams. The vocabulary that was learned shows how people really use it in Tarifit user-generated content, such as code-switching, emphasis on repetition, hybrid loanwords, and emoji semantics.

The model achieves excellent results in both model accuracy and data prediction ability. The model gets almost perfect accuracy on the training set (0.9906) and stays very good at generalizing to new data, getting 0.8800 accuracy and 0.8798 Macro-F1 on the test set. The predictive behavior maintains equal distribution between all sentiment categories according to the class-wise analysis. The model produces 0.8486 precision for negative class predictions, while it shows excellent recall performance at 0.9250 and achieves an F1-score of 0.8852 across 200 instances. The system shows high sensitivity to negative polarity indicators, which it uses for its operations. The positive class maintains its performance at a high level because it achieves precision at 0.9176, recall at 0.8350, and F1-score at 0.8743 across 200 examples. The results demonstrated that positive sentiment received correct predictions at suitable levels, which stayed below the predicted thresholds. The results show that the proposed method maintains consistent performance, which remains balanced between all sentiment categories while achieving good generalization capabilities.

B. Comparison with Other Models

1) Static lexicon baseline: The static lexicon baseline is the simplest modeling strategy and is used as a lower-bound reference. It only uses direct dictionary lookups [26] to assign polarity based on a manually curated lexicon of 200 entries (100 positive and 100 negative). The system fails to detect language patterns in specific contexts because it does not understand how words relate to each other when used in different situations. The model doesn't work well, though, with an accuracy of only 0.5275 and a Macro-F1 of 0.3917. This shows a strong class bias and a failure to capture balanced sentiment behavior. The confusion matrix also shows that almost all instances, including many negative comments, are misclassified as positive. In general, this baseline doesn't show any sensitivity to important sentiment phenomena that are common in Tarifit discourse, like negation, spelling variation, mixed-script writing (Arabic, Latin, and Tifinagh), expressive elongation, or emoji-based sentiment cues.

2) TF–IDF + SVM baseline: TF-IDF and a Support Vector Machine together make a strong and widely used traditional machine-learning baseline for classifying text [27]. This model uses statistical term weighting to find discriminative lexical patterns directly from data, which is different from lexicon-based approaches that need pre-defined linguistic rules or sentiment inventories. In our tests, TF–IDF + SVM got 0.8525 accuracy and 0.8512 Macro-F1, showing that it can handle sparse and high-dimensional text representations that are common in social media data. The corresponding confusion matrices for D-LexeCan, the static lexicon baseline, and TF–IDF + SVM on the test set are presented in Fig. 3.

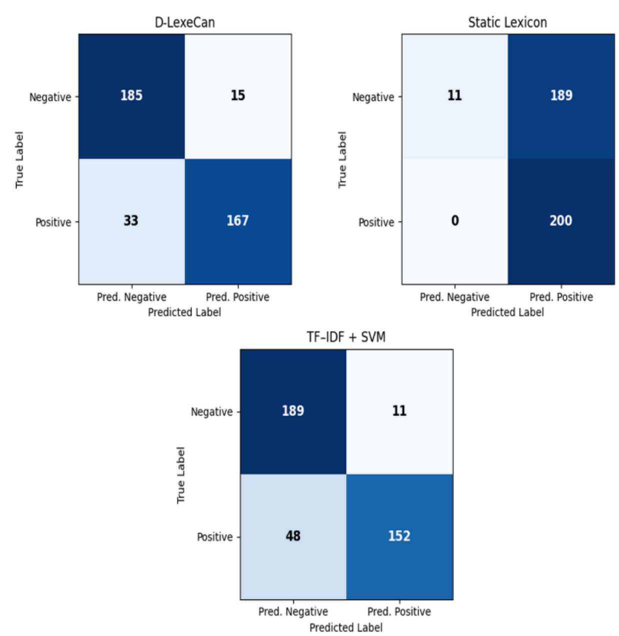

Fig. 3. Confusion matrices on the test set for D-LexeCan, static lexicon, and TF–IDF + SVM.

These results show that the model can accurately separate sentiment classes just by looking at the distributional cues. Despite its strong ability to make predictions, this method has two major problems: first, its decision-making process is mostly unclear, making it harder to understand than lexicon-based methods; second, it doesn't have a built-in way to quickly add new sentiment expressions or vocabulary, which makes it hard to keep up with quickly changing informal language without retraining on new data.

3) Frozen transformer embeddings: Multilingual transformer models that have been trained on large, diverse datasets are often used as general-purpose text encoders for cross-lingual tasks. This is because they give contextual representations without needing to be fine-tuned for each task [28]. The models were first tested as fixed encoders, which worked with conventional classification systems for Tarifit sentiment analysis. The results show only moderate performance, with test accuracy between 0.6725 and 0.7650. The research findings would not provide useful results because the language shows limited orthographic diversity, extensive morphological complexity, and unique sentiment expression patterns. The XLM-RoBERTa + SVM configuration has the best accuracy (0.7650, Macro-F1 = 0.7649), followed closely by mBERT + Logistic Regression (0.7625, Macro-F1 = 0.7620) and mBERT + SVM (0.7525, Macro-F1 = 0.7518). mBERT + KNN (0.7150, Macro-F1 = 0.7136) does worse than mBERT + Naive Bayes (0.6725 accuracy, Macro-F1 = 0.6683). In general, these results show that frozen multilingual transformers can capture a lot of semantic information, but they are not good enough for fine-grained sentiment modeling in Tarifit. The results indicate that these models require modifications to reach successful performance in particular tasks, which require specific language applications.

4) Fine-tuned transformer models: Fine-tuning multilingual transformer models lets their internal representations be changed directly to match sentiment-related patterns in the target language. The model becomes more effective at detecting task-related language patterns and contextual information, and polarity signals through this approach. This method makes the model more sensitive to sentiment expressions that aren't well represented in generic pretrained embeddings by changing the model parameters on labeled data [29]. In our tests, fine-tuned mBERT gets 0.8175 accuracy and 0.8170 Macro-F1, which shows that it can model sentiment in Tarifit very well after being adapted to the task. Fine-tuned XLM-RoBERTa gets 0.7650 accuracy and 0.7646 Macro-F1, which shows that fine-tuning doesn't help it much. The comparative test accuracy distribution of frozen and fine-tuned transformer models is summarized in Fig. 4. The research demonstrates that fine-tuning enhances sentiment modeling through better target task adaptation of pre-trained models, although it requires additional computational resources and creates training complexities.

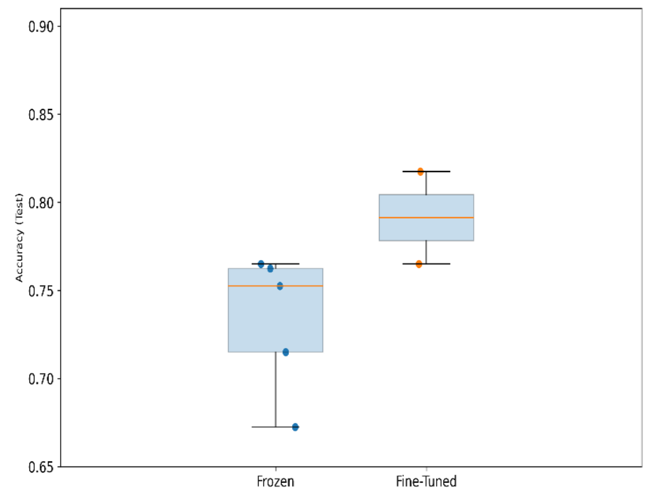

Fig. 4. Test accuracy distribution of transformer models (frozen vs. fine-tuned).

5) BiLSTM (End-to-End): The BiLSTM network functions as a popular neural sequence model for sentiment analysis because it analyzes text sequences from both start to end and end to start, which reveals the sequence of words and their contextual relationships [30]. The training process of BiLSTMs allows them to create task-oriented representations directly from labeled data without requiring any pre-trained linguistic resources. When used on normalized Tarifit text, the BiLSTM model gets 0.7950 accuracy and 0.7903 Macro-F1, which shows that it can reasonably pick up on sequential sentiment cues. These results show how good recurrent architectures are at modeling contextual information and how hard it is for purely data-driven neural models to work in low-resource settings where there isn't much training data and there is a lot of linguistic variation.

6) Overall comparison of models: The evaluation of performance and robustness needed different sentiment classification methods, which included symbolic lexicon-based approaches, machine-learning models, neural architectures, and transformer-based frameworks. All models received identical experimental conditions during training and testing, which allowed for an equal comparison between them. The performance results from the comparison appear in Fig. 5, which shows how different modeling approaches performed relative to each other.

C. Discussion

The experimental results show that dynamic lexicon induction presents a successful approach for sentiment analysis, which uses linguistic principles to analyze three types of languages, including Tarifit, which has limited resources and uses informal language and different writing systems. The static lexicon baseline achieved 0.5275 accuracy and 0.3917 Macro-F1, but the proposed dynamic D-Lexicon within D-LexeCan outperformed it by achieving 0.8800 accuracy and 0.8798 Macro-F1. The results show that static sentiment inventories face essential limitations when used for processing user-generated content, which contains script mixing and dialectal variation, expressive elongation, and emoji usage and lexical changes over time.

The research results show that lexical adaptability stands as a crucial element that affects results regardless of the model complexity level. The static lexicon has two main restrictions because it includes only 200 entries and does not have any system to add new words that carry sentiment or to adjust emotional strength through actual data or to identify words which require contextual understanding. The dynamic lexicon induction strategy used in D-LexeCan, on the other hand, lets the model learn polarity information directly from annotated data. The model can learn to accept dialectal variations and non-standard writing and emoji meanings and typical multi-word phrases through this approach. The lexicon size (9,325 entries, including 3,974 unigrams and 5,351 bigrams) shows how many different ways people express their feelings on Tarifit social media.

The system, which combines TF-IDF with SVM, produces better results than conventional machine learning methods (0.8525 accuracy / 0.8512 Macro-F1). This confirms that it is a good baseline for low-resource sentiment analysis. But this method depends on high-dimensional feature representations, which makes it hard to understand and inspect, reuse, or add to the learned sentiment knowledge. D-LexeCan, on the other hand, makes an explicit lexicon that lets you look at and change sentiment entries and their scores as more labeled data becomes available. The system enables users to access a simplified system that helps them study and assess language content.

Neural and transformer-based models demonstrate multiple advantages and disadvantages, which create an effective partnership between them. The frozen multilingual transformer models achieved average performance results through their accuracy range of 0.6725 to 0.7650. This suggests that generic pretrained representations may not fully match the orthographic variability and sentiment conventions of Tarifit. The system performance improves through fine-tuning because mBERT produces the highest results at 0.8175 accuracy and 0.8170 Macro-F1, which demonstrates its capability to execute multiple tasks successfully. The system has received improved features, but it requires more complex processing operations while its operation remains difficult to understand. The neural representations contain sentiment-related information about negation scope and amplification effects, and emoji polarity and multiscript variation, which makes their direct analysis more challenging.

The results from D-LexeCan show important value because this system operates at a high level while using minimal resources and maintaining easy-to-understand operations. The model recognizes patterns in sentiment data instead of learning instances from memory because it achieves 0.9906 training accuracy, 0.8800 accuracy, and 0.8798 Macro-F1 performance on the test set. The results from class-wise evaluation support this analysis because the model shows excellent ability to detect negative sentiment with a recall of 0.9250, and it achieves high precision for positive sentiment at 0.9176, which results in equal performance for all sentiment categories Fig. 5.

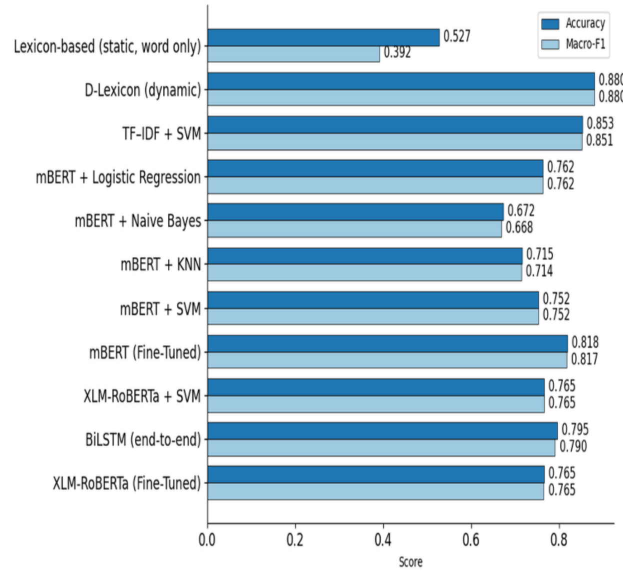

Fig. 5. Overall performance comparison across sentiment classification models.

In general, these results support the idea that lexicon-based frameworks that are based on language and data can be a good addition to neural approaches for sentiment analysis in low-resource settings. The research focuses on Tarifit binary sentiment classification, but the established design principles for multiscript normalization, confidence-based polarity induction, and bigram-level pattern learning and operator-aware scoring could help solve similar challenges that other low-resource languages face. The future of research should focus on three main areas, which include multi-class sentiment analysis, cross-domain evaluation, and online lexicon update systems to determine how well dynamic lexicon-based methods work across different situations.

V. Conclusion

The research presented D-LexeCan as a dynamic framework that provides interpretable results through linguistic analysis for Tarifit (Tamazight of the Rif) sentiment analysis of this severely underdocumented language, which uses multiple scripts and shows unstable writing systems and contains mostly casual user-generated content. The proposed framework solves two problems that static lexicons cannot identify and that generic pretrained architectures detect only through hidden mechanisms. The system achieves this through three main components, which include deterministic multiscript normalization into Tarifit Latin representation and corpus-driven induction of sentiment-bearing unigrams and bigrams, and explicit modeling of linguistic operators, including amplification and negation patterns and emoji cues. The experimental assessment demonstrates that D-LexeCan attains strong and competitive performance in comparison to traditional machine-learning benchmarks, as well as neural and transformer-based models, while ensuring transparency and computational efficiency. This work highlights that the proposed framework has broader implications for low-resource NLP, as it shows that dynamic lexicon induction can be an effective and interpretable alternative when annotated data is scarce. Although evaluated on Tarifit, the approach is generalizable to other low-resource, multi-dialectal, and multi-script languages with similar linguistic challenges. The proposed framework enables users to apply their learned lexical knowledge for sentiment analysis, which they can confirm and adjust for upcoming applications. The system operates through a different method than fine-tuned multilingual transformers because it needs task-specific parameter adaptation, and it needs more computing power. The research findings show that dynamic lexicon induction with explicit linguistic modeling produces the best solution for sentiment analysis, which performs better than neural methods when operating under limited resource conditions. The Tarifit case demonstrates how D-LexeCan script-agnostic normalization and confidence-based polarity induction, multiword sentiment modeling, and operator-aware scoring methods can help solve structural and sociolinguistic problems in under-resourced languages resulting in better sentiment analysis in multilingual environments.

REFERENCES

[1] D. Karamouzas, I. Mademlis, et I. Pitas, « Public opinion monitoring through collective semantic analysis of tweets », Soc. Netw. Anal. Min., vol. 12, no 1, p. 91, déc. 2022, doi: 10.1007/s13278-022-00922-8.

[2] A. Hassani et E. Mosconi, « Social media analytics, competitive intelligence, and dynamic capabilities in manufacturing SMEs », Technological Forecasting and Social Change, vol. 175, p. 121416, 2022.

[3] N. Nasrabadi, H. Wicaksono, et O. F. Valilai, « Shopping marketplace analysis based on customer insights using social media analytics », MethodsX, vol. 13, p. 102868, 2024.

[4] R. Menaha et K. Ananthi, « Reviewing the effectiveness of lexicon-based techniques for sentiment analysis in massive open online courses », Int J Data Sci Anal, vol. 20, no 3, p. 1631-1642, sept. 2025, doi: 10.1007/s41060-024-00585-y.

[5] O. Alsemaree, A. S. Alam, S. S. Gill, et S. Uhlig, « An analysis of customer perception using lexicon-based sentiment analysis of Arabic Texts framework », Heliyon, vol. 10, no 11, 2024, Consulté le: 22 décembre 2025. [En ligne]. Disponible sur: https://www.cell.com/heliyon/fulltext/S2405-8440(24)06351-5

[6] M. Rodríguez-Ibánez, A. Casánez-Ventura, F. Castejón-Mateos, et P.-M. Cuenca-Jiménez, « A review on sentiment analysis from social media platforms », Expert Systems with Applications, vol. 223, p. 119862, 2023.

[7] T. A. Al-Qablan, M. H. Mohd Noor, M. A. Al-Betar, et A. T. Khader, « A survey on sentiment analysis and its applications », Neural Comput & Applic, vol. 35, no 29, p. 21567-21601, oct. 2023, doi: 10.1007/s00521-023-08941-y.

[8] Y. Aliyu, A. Sarlan, K. U. Danyaro, A. S. B. Rahman, et M. Abdullahi, « Sentiment analysis in low-resource settings: a comprehensive review of approaches, languages, and data sources », IEEE Access, vol. 12, p. 66883-66909, 2024.

[9] N. Raychawdhary, A. Das, S. Bhattacharya, G. Dozier, et C. D. Seals, « Optimizing multilingual sentiment analysis in low-resource languages with adaptive pretraining and strategic language selection », in 2024 IEEE 3rd International Conference on Computing and Machine Intelligence (ICMI), IEEE, 2024, p. 1-5. Consulté le: 22 décembre 2025. [En ligne]. Disponible sur: https://ieeexplore.ieee.org/abstract/document/10585876/

[10] N. Raychawdhary, S. Bhattacharya, C. Seals, et G. Dozier, « Empowering Sentiment Analysis in African Low-Resource Languages through Transformer Models and Strategic Language Selection », IEEE Access, 2025, Consulté le: 22 décembre 2025. [En ligne]. Disponible sur: https://ieeexplore.ieee.org/abstract/document/11127074/

[11] T. Amzil, A. Jannani, et N. Sael, « Sentiment Analysis for Low-Resource and Arabic Dialects: A Comprehensive Review », in 2025 International Conference on Circuit, Systems and Communication (ICCSC), IEEE, 2025, p. 1-7. Consulté le: 22 décembre 2025. [En ligne]. Disponible sur: https://ieeexplore.ieee.org/abstract/document/11135025/

[12] S. El Ouahabi, S. El Ouahabi, et E. W. Dadi, « Contribution to the Moroccan Darija sentiment analysis in social networks », Soc. Netw. Anal. Min., vol. 13, no 1, p. 138, oct. 2023, doi: 10.1007/s13278-023-01129-1.

[13] N. Mechulam, D. Salvia, A. Rosá, et M. Etcheverry, « Building dynamic lexicons for sentiment analysis », Inteligencia Artificial, vol. 22, no 64, p. 1-13, 2019.

[14] V. R. N. S. S. V. S. P, « Dynamic Lexicon-Based Sentiment Analysis Architecture Using Nonlinear Feature Optimization », Communications on Applied Nonlinear Analysis, vol. 32, no 1s, p. 557-572, 2025, doi: 10.52783/cana.v32.2344.

[15] F. Koto, T. Beck, Z. Talat, I. Gurevych, et T. Baldwin, « Zero-shot Sentiment Analysis in Low-Resource Languages Using a Multilingual Sentiment Lexicon », 3 février 2024, arXiv: arXiv:2402.02113. doi: 10.48550/arXiv.2402.02113.

[16] K. M. Awlla, H. Veisi, et A. A. Abdullah, « Sentiment analysis in low-resource contexts: BERT's impact on Central Kurdish », Lang Resources & Evaluation, vol. 59, no 3, p. 2213-2243, sept. 2025, doi: 10.1007/s10579-024-09805-0.

[17] Y. Pei, S. Chen, Z. Ke, W. Silamu, et Q. Guo, « Ab-labse: Uyghur sentiment analysis via the pre-training model with bilstm », Applied Sciences, vol. 12, no 3, p. 1182, 2022.

[18] A. Ali, M. Khan, K. Khan, R. U. Khan, et A. Aloraini, « Sentiment Analysis of Low-Resource Language Literature Using Data Processing and Deep Learning. », Computers, Materials & Continua, vol. 79, no 1, 2024, Consulté le: 22 décembre 2025. [En ligne]. Disponible sur: https://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=15462218&AN=176916269&h=DMsPLzYT6m5KBw3K6EoPUZLWKMFzr1a8JUZrzSsUV2LcrtYLhlh5nALksjpORpeeeiaD230o%2Fyp962GaFR8omw%3D%3D&crl=c

[19] S. Iqbal, F. Khan, H. U. Khan, T. Iqbal, et J. H. Shah, « Sentiment analysis of social media content in pashto language using deep learning algorithms », Journal of Internet Technology, vol. 23, no 7, p. 1669-1677, 2022.

[20] L. S. Meetei, T. D. Singh, S. K. Borgohain, et S. Bandyopadhyay, « Low resource language specific pre-processing and features for sentiment analysis task », Lang Resources & Evaluation, vol. 55, no 4, p. 947-969, déc. 2021, doi: 10.1007/s10579-021-09541-9.

[21] P. Liu, C. Marco, et J. A. Gulla, « Semi-supervised Sentiment Analysis for Under-Resourced Languages with a Sentiment Lexicon. », in INRA@ RecSys, 2019, p. 12-17. Consulté le: 22 décembre 2025. [En ligne]. Disponible sur: https://ceur-ws.org/Vol-2554/paper_02.pdf

[22] V. Dhananjaya, S. Ranathunga, et S. Jayasena, « Lexicon‐based fine‐tuning of multilingual language models for low‐resource language sentiment analysis », CAAI Trans on Intel Tech, vol. 9, no 5, p. 1116-1125, oct. 2024, doi: 10.1049/cit2.12333.

[23] I. Mohammed et R. Prasad, « Building lexicon-based sentiment analysis model for low-resource languages », MethodsX, vol. 11, p. 102460, 2023.

[24] M. Bordoloi et S. K. Biswas, « Sentiment analysis: A survey on design framework, applications and future scopes », Artif Intell Rev, vol. 56, no 11, p. 12505-12560, nov. 2023, doi: 10.1007/s10462-023-10442-2.

[25] M. Cherradi et A. El Haddadi, « Comparative analysis of machine learning algorithms for sentiment analysis in film reviews », Acadlore Trans. Mach. Learn, vol. 3, no 3, p. 137-147, 2024.

[26] A. M. van der Veen et E. Bleich, « The advantages of lexicon-based sentiment analysis in an age of machine learning », PloS one, vol. 20, no 1, p. e0313092, 2025.

[27] D. E. Cahyani et I. Patasik, « Performance comparison of tf-idf and word2vec models for emotion text classification », Bulletin of Electrical Engineering and Informatics, vol. 10, no 5, p. 2780-2788, 2021.

[28] S. T. Kokab, S. Asghar, et S. Naz, « Transformer-based deep learning models for the sentiment analysis of social media data », Array, vol. 14, p. 100157, 2022.

[29] M. P. Geetha et D. K. Renuka, « Improving the performance of aspect based sentiment analysis using fine-tuned Bert Base Uncased model », International Journal of Intelligent Networks, vol. 2, p. 64-69, 2021.

[30] A. J. Lak, R. Boostani, F. A. Alenizi, A. S. Mohammed, et S. M. Fakhrahmad, « RoBERTa, ResNeXt and BiLSTM with self-attention: The ultimate trio for customer sentiment analysis », Applied Soft Computing, vol. 164, p. 112018, 2024.