

Original Research

Received 09 June 2026

Revised 20 July 2026

Accepted 03 Aug 2026

Natural Resources for Human Health 2026, Vol. 6 (13s): 770–788

KEYWORDS:

Machine Learning; Natural Language Processing; Computational Literary Studies; Character Emotion; Psychological States; Emotion Detection; Literary Narratives; Transformer Models; BERT; Digital Humanities; Affective Computing; Sentiment Analysis.

Machine Learning-Based Detection of Character Emotions and Psychological States in Literary Narratives

¹Dr. Vandana Singh, ²Dr. Abhilasha Singh, ³Dr. Brijesh Verma, ⁴Prof. Bhavana Shirke, ⁵Dr. Sanju Harne, ⁶Dr. Sourabh Dubey

¹Department of Humanities, Lakshmi Narain College of Technology Excellence, Bhopal, Madhya Pradesh

²Department of Humanities, Lakshmi Narain College of Technology & Science Bhopal, Madhya Pradesh

³Department of Humanities, Lakshmi Narain College of Technology Excellence, Bhopal, Madhya Pradesh

⁴Department of Humanities, Bansal Institute of Science and Technology, Bhopal, Madhya Pradesh

⁵Department of Humanities, Bansal Institute of Science and Technology, Bhopal, Madhya Pradesh

⁶Nobel Public Higher Secondary School, Mandsaur, Madhya Pradesh

1vchauhann@gmail.com, 2abhilashas@lnct.ac.in,

3brijeshv@lnct.ac.in, 4bhavanashirke30@gmail.com,

5sh.bist.24@gmail.com, 6dubey0130@gmail.com

Corresponding Author: Vandana Singh, vchauhann@gmail.com

ABSTRACT: The representation of emotions and psychological states is central to literary narratives, yet identifying such states systematically across large collections of fiction remains a challenging task for traditional literary criticism. Recent developments in Natural Language Processing (NLP), machine learning (ML), deep learning, and transformer-based language models provide new possibilities for computationally analyzing character emotions, affective transitions, and psychologically significant narrative patterns. This study proposes a machine-learning-based framework for detecting and analyzing character emotions and psychological states in English literary narratives. The proposed framework integrates character identification, co-reference resolution, dialogue attribution, contextual emotion classification, sentiment analysis, psychological-state modeling, and temporal tracking of emotional trajectories. Unlike conventional sentiment analysis, which generally classifies the emotional polarity of a text segment, the proposed approach attempts to identify who experiences an emotion, what emotion is expressed, in what narrative context, and how that emotional state changes over time. The proposed methodology treats psychological-state detection as a computationally assisted interpretive task rather than as clinical diagnosis. It distinguishes textual evidence of states such as fear, grief, anxiety, loneliness, hope, anger, guilt, and emotional conflict from medical or psychiatric diagnosis. A research design is presented using annotated literary passages, transformer-based classification; explainability techniques, and character-level temporal modeling. Illustrative examples demonstrate how the framework can identify emotional transitions in fictional characters. The paper contributes a multidisciplinary bridge between computational

literary studies, NLP, machine learning, and affective literary criticism and provides a reproducible framework for future empirical research.

I. INTRODUCTION

Emotion is one of the fundamental components of literary narrative. Characters experience fear, love, anger, grief, jealousy, hope, guilt, loneliness, excitement, disappointment, and numerous complex combinations of emotional states. These emotions influence character development, interpersonal relationships, narrative conflict, plot progression, and reader engagement. Literary scholars have traditionally examined these dimensions through close reading, narratology, psychoanalytic criticism, reader-response theory, stylistics, and other interpretive approaches. However, the increasing availability of digitized literary texts has created opportunities for computational approaches to literary analysis. A researcher can now examine hundreds or thousands of novels using computational methods that would be extremely difficult to analyze manually. Machine learning and NLP can identify linguistic patterns associated with emotions, characters, events, relationships, and narrative structures.

The computational study of emotion has developed rapidly in NLP. Early approaches frequently depended on sentiment dictionaries and manually constructed lexical resources. The NRC Emotion Lexicon, for example, associates words with eight emotions—anger, anticipation, disgust, fear, joy, sadness, surprise, and trust—and positive/negative sentiment. (National Research Council Canada)[1] Such resources remain valuable because they provide interpretable lexical evidence. Nevertheless, literary language presents difficulties for purely lexical systems because the emotional meaning of a word depends heavily on context.

Consider the sentence:

“She smiled as she watched the train disappear.”

The word *smiled* may suggest happiness. However, in a particular narrative context, the smile could represent resignation, sadness, irony, or emotional suppression. Similarly:

“He was calm.”

May describe genuine emotional stability, concealed anger, emotional exhaustion, or deliberate self-control depending on surrounding passages. These examples demonstrate why literary emotion detection cannot be reduced to counting positive and negative words. Recent transformer-based models have significantly improved contextual language representation. BERT and related architectures can represent words according to their surrounding linguistic context. Large emotion datasets have further encouraged fine-grained classification. Go Emotions, for example, provides 27 emotion categories plus neutral and contains approximately 58,000 manually annotated English comments. Its authors report that their BERT-based baseline achieved an average F1 score of .46 across the proposed taxonomy, illustrating both the potential and the continuing difficulty of fine-grained emotion classification. Literary narratives create an additional challenge: emotion belongs to characters, not merely sentences. A narrative sentence may describe one character observing another character's emotion. Dialogue may contain an emotion expressed by a speaker but attributed by the narrator to another person. Pronouns and anaphoric expressions may obscure the identity of the emotional experience. Research specifically addressing affect analysis in narratives has therefore highlighted character/person identification and anaphora as important components of the problem.[2]

The present paper proposes a computational framework addressing this challenge. Rather than treating literary text as an undifferentiated collection of emotional sentences, the proposed model constructs a **character-emotion representation**.

The central research problem is therefore:

How can machine-learning techniques identify, classify, and temporally track emotions and psychologically significant states associated with individual characters in literary narratives?

The study contributes to this problem by integrating:

1. Character Recognition;
2. Co reference Resolution;
3. Dialogue Attribution;
4. Contextual Emotion Classification;
5. Sentiment Analysis;
6. Psychological-State Modeling;
7. Character-Level Temporal Tracking; And
8. Interpretable Literary Analysis.

II. BACKGROUND AND CONCEPTUAL FRAMEWORK

i. Computational Literary Studies

Computational literary studies apply computational methods to literary texts to investigate patterns that may be difficult to identify through conventional close reading alone. It does not necessarily replace human interpretation. Instead, computational analysis can function as a complementary methodology. A literary scholar may identify a character's increasing isolation through close reading. A computational model can potentially examine whether linguistic indicators of sadness, fear, loneliness, and social withdrawal increase across hundreds of passages associated with that character.

This creates a productive relationship between quantitative and qualitative interpretation:

Machine learning → pattern detection → human interpretation → literary meaning

The proposed research follows this human-in-the-loop model.

ii. Emotion and Sentiment

Sentiment analysis generally distinguishes positive, negative, and sometimes neutral attitudes. Emotion detection is more fine-grained.

For example:

Textual State	Possible classification
"She laughed with delight."	Joy
"He trembled at the sound."	Fear
"She could not forgive him."	Anger/resentment
"He mourned silently."	Grief/sadness
"She hoped that he would return."	Hope/optimism
"He regretted what he had done."	Remorse/guilt
"She did not know what to choose."	Confusion
"He wanted to leave."	Desire

The distinction is important because two passages may both be classified as negative sentiment while representing very different psychological experiences. The NRC resource is particularly useful for lexical emotion analysis, but researchers have also identified limitations and contextual biases in emotion lexicons. For example, subsequent research has documented

problematic entries in the NRC lexicon and argued for contextual and part-of-speech-sensitive correction. Consequently, the proposed framework uses lexical resources as supporting evidence, rather than treating them as final authorities. [3]

III. LITERATURE REVIEW

i. Emotion Detection in NLP

Emotion detection has traditionally been approached through lexicon-based, machine-learning, and deep-learning techniques. Lexicon-based systems associate words with emotions [4]. The NRC Emotion Lexicon represents one important resource, containing associations between words and eight basic emotions. Its interpretability makes it particularly useful in digital humanities research where researchers may want to explain why a computational system assigned an emotion.

Traditional machine-learning systems use features such as:

- Word Frequency;
- TF-IDF;
- N-Grams;
- Part-Of-Speech Patterns;
- Syntactic Features;
- Sentiment Scores;
- Lexical Emotion Features.

Common classifiers include:

- Naïve Bays;
- Logistic Regression;
- Support Vector Machines;
- Random Forest;
- Decision Trees.

Deep-learning approaches subsequently introduced CNNs, recurrent neural networks, LSTMs, and BILSTMS. Transformer architectures have further changed NLP by allowing contextual representations to capture relationships among words across a sequence. BERT, introduced by Devlin et al. (2019)[5] demonstrated the effectiveness of bidirectional transformer representations for multiple language understanding tasks.

ii. Fine-Grained Emotion Classification

A major development in emotion research is the movement from basic emotion categories toward fine-grained taxonomies. Go Emotions is particularly relevant because it contains 27 emotion categories and a neutral category.[6] The dataset includes categories such as admiration, amusement, anger, annoyance, approval, caring, confusion, curiosity, desire, disappointment, disapproval, disgust, embarrassment, excitement, fear, gratitude, grief, joy, love, nervousness, optimism, pride, realization, relief, remorse, and sadness. Its importance for the present research lies in demonstrating that textual emotion can be represented through a richer taxonomy than simple positive/negative sentiment.

However, Go Emotions was constructed from Reddit comments rather than literary fiction. Domain transfer therefore becomes a significant methodological issue. Literary narratives contains[7]:

- Archaic Vocabulary;
- Metaphor;

- Indirect Emotional Expression;
- Free Indirect Discourse;
- Narrative Description;
- Dialogue;
- Irony;
- Symbolism;
- Unreliable Narration.

Consequently, a model trained on conversational data cannot automatically be assumed to perform equally well on novels.

iii. Emotion Detection in Narratives

Narrative texts create a special challenge because multiple characters may appear in the same sentence or paragraph. Research on affect analysis in narratives has explicitly identified the difficulty of determining the emotional subject in sentences containing different characters and narrative forms. One approach has used anaphoric analysis to identify the subject of an emotional expression. For example:

“John looked at Mary. She was trembling.”

A sentiment model may identify *trembling* as fear-related. But the system must establish that **Mary**, rather than John, has the experience.

A second example is:

“John saw Mary crying and felt helpless.”

Here the text contains two affective states:

- Mary → Sadness/Grief;
- John → Helplessness/Distress.

Character-level emotion detection therefore requires entity tracking and semantic role analysis.

iv. AI and Psychological-State Detection in Literature

Recent research has begun applying AI and NLP to psychological conditions represented in literary narratives. A 2023 study, for example, examined depression-related representations in selected Afghan and Pakistani diasporic novels using topic modeling, sentiment analysis, and machine-learning classifiers. The study reported the use of LDA, LSI, HDP, NMF, VADER, Text Blob, Decision Trees, Naïve Bays, KNN, and SVM[8].

This work demonstrates the feasibility of computational approaches to psychologically significant literary themes. However, psychological-state detection in literature must be distinguished from clinical diagnosis. A fictional character is a textual construct. A computational model can identify linguistic indicators associated with anxiety, grief, isolation, fear, or depressive themes, but such identification does not establish a clinical disorder.

This distinction is fundamental to the proposed framework.

IV. OBJECTIVES OF THE STUDY

The study has the following objectives:

1. To develop a machine-learning framework for detecting emotions associated with literary characters.
2. To distinguish character-level emotions from general narrative sentiment.

3. To classify fine-grained emotions using contextual NLP techniques.
4. To identify textual indicators of psychologically significant states.
5. To model emotional changes across narrative time.
6. To compare lexical and transformer-based approaches.
7. To develop an interpretable framework suitable for literary scholars.
8. To identify methodological limitations associated with automated psychological-state detection.

V. METHODOLOGY

i. Research Design

The study proposes a mixed computational and interpretive research design consisting of five major stages:

Corpus preparation → Character identification → Emotion classification → Psychological-State Modeling → Literary Interpretation[20]

The methodology can be implemented using Python-based NLP libraries and transformer architectures.

ii. Corpus Selection

The recommended corpus should contain digitized English literary narratives from multiple genres.

Possible categories include:

- Victorian Fiction;
- Modern Fiction;
- Contemporary Fiction;
- Gothic Fiction;
- Psychological Fiction;
- Diasporic Literature;
- Postcolonial Fiction.

A balanced corpus could contain approximately 20–50 novels for an initial study, with additional texts reserved for external validation.

The corpus should ideally contain:

- Complete Texts;
- Identifiable Chapters;
- Stable Paragraph Boundaries;
- Character Names;
- Metadata Concerning Author, Period, and Genre.

Copyright restrictions must be respected. Public-domain texts are particularly suitable for reproducible research.

VI. DATA PRE-PROCESSING

The following processing pipeline is proposed:

Step-1: Text Cleaning

Remove:

- OCR Errors;
- Page Numbers;
- Publisher Information;
- Repeated Headers;
- Irrelevant Metadata.

Step-2: Sentence Segmentation

The text is divided into sentences.

Step-3: Tokenization

Each sentence is converted into tokens.

Step-4: Part-of-Speech Tagging

Words are assigned grammatical categories.

Step-5: Named Entity Recognition

Character names and other entities are identified.

Step-6: Co reference Resolution

Pronouns such as:

- He;
- She;
- They;
- Him;
- Her;
- His;
- Their

Are linked to the appropriate character entities.

Step-7: Dialogue identification

Quoted speech is separated from narration.

Step-8: Character-event linking

Emotionally relevant events are linked to characters.

VII. CHARACTER IDENTIFICATION

Character identification is a fundamental component of the framework.

Consider:

“Elizabeth entered the room. She looked frightened.”

A basic emotion classifier may identify *frightened*^[9] as fear. A character-aware system additionally establishes:

Elizabeth → fear

The system therefore represents an emotional event as:

$E = (C, T, Em, S, P)$

Where:

- C = Character;
- T = Textual Location;
- Em = Emotion;
- S = Intensity Score;
- P = Evidence Passage

This representation permits character-level analysis.

VIII. EMOTION CLASSIFICATION MODEL

The proposed framework combines three levels.

i. Lexicon-Based Layer

The NRC Emotion Lexicon provides interpretable lexical features. Its eight emotion categories include:

1. Anger;
2. Anticipation;
3. Disgust;
4. Fear;
5. Joy;
6. Sadness;
7. Surprise;
8. Trust.

The resource is useful for identifying explicit emotion-related vocabulary.

However, lexical classification alone cannot reliably resolve context.

ii. Traditional Machine-Learning Layer

TF-IDF (Term Frequency–Inverse Document Frequency)[10] features can be combined with:

- Logistic Regression;
- SVM;
- Random Forest;
- Naïve Bays.

A simple **TF-IDF (Term Frequency–Inverse Document Frequency)** representation can be represented mathematically as:

$$TF-IDF(t,d)=TF(t,d)* IDF(t)$$

Where:

- $TF(t,d)$ = Term Frequency of term t in document d

$$TF(t,d) = (\text{Number of occurrences of } t \text{ in } d) / (\text{Total number of terms in } d)$$

- $IDF(t)$ = Inverse Document Frequency of term t

$$IDF(t) = \log(N / (1+DF(t)))$$

Where:

- N = Total number of documents
- $DF(t)$ = Number of documents containing term t .

Thus, a document can be represented as a **TF-IDF vector**:

$$d=[w_1,w_2,w_3,\dots,w_V]$$

Where, each w_i is the TF-IDF weight of the i^{th} term in the vocabulary of size V .

These models provide useful baselines.

iii. Transformer Layer

A transformer-based model such as BERT can be fine-tuned for emotion classification.

For an input sequence:

$$X = (x_1, x_2, \dots, x_n)$$

$$X = (x_{-1}, x_{-2}, \dots, x_{-n})$$

The transformer produces contextual representations:

$$H = (h_1, h_2, \dots, h_n)$$

$$H = (h_{-1}, h_{-2}, \dots, h_{-n})$$

A classification layer then estimates:

$$P(E|X) = \text{SOFTMAX}(WHCLS+B)$$

$$P(E|X) = \text{SOFTMAX}(Wh_{\{CLS\}}+b)$$

Where, $WHCLS$ and $Wh_{\{CLS\}}$ represent the contextual sequence representations.

For literary analysis, the input should ideally contain surrounding context rather than only the sentence containing an emotional word.

For example:

Previous context + target sentence + following context

Can be supplied to the classifier.

IX. FINE-GRAINED EMOTION TAXONOMY

The proposed framework adopts a hybrid taxonomy based on established emotion resources and literary requirements.

Category	Examples
Joy	Happiness, Delight, Pleasure
Sadness	Sorrow, Grief, Despair
Fear	Terror, Anxiety, Apprehension
Anger	Rage, Irritation, Resentment
Love	Affection, Attachment
Guilt	Remorse, Self-Blame
Shame	Humiliation, Embarrassment
Hope	Optimism, Expectation
Loneliness	Isolation, Abandonment
Confusion	Uncertainty, Perplexity
Desire	Longing, Craving
Surprise	Astonishment, Shock
Trust	Confidence, Reliance
Disgust	Revulsion, Contempt
Relief	Release From Fear/Stress
Emotional conflict	Simultaneous or Competing States

Go Emotions demonstrates the usefulness of fine-grained emotion categories, although its original domain is conversational rather than literary. (ACL Anthology)[11]

X. PSYCHOLOGICAL-STATE MODELING

Psychological-state detection should operate at the level of **textual indicators**, rather than clinical diagnosis.

For example:

Textual indicators	Possible literary state
Repeated Expressions of Isolation	Loneliness
Persistent Fear-Related Language	Anxiety/Fear
References to Loss and Mourning	Grief
Self-Blaming Language	Guilt
Hopelessness and Withdrawal	Despair/Depressive Representation
Anticipation of Future Events	Hope
Conflicting Emotional Statements	Psychological Conflict
Inability to Make Decisions	Uncertainty/Confusion

The model should therefore output:

“Textual evidence consistent with grief”

Rather Than:

“Character has a clinical depressive disorder.”

This distinction protects interpretive validity and avoids inappropriate clinical inference.

XI. CHARACTER EMOTION VECTOR

A character's emotional state at time t can be represented as:

$C_t = [J, S, F, A, L, G, H, D, R]$

$C_{_t} = [J, S, F, A, L, G, H, D, R]$

Where:

- J = Joy;
- S = Sadness;
- F = Fear;
- A = Anger;
- L = Love;
- G = Guilt;
- H = Hope;
- D = Desire;
- R = Relief.

For each chapter:

$C_t = f(E_t)$

$C_{_t} = f(E_{_t})$

Where, E_t and $E_{_t}$ represent the detected emotional events associated with the characters.

The resulting sequence:

$C_1, C_2, C_3, \dots, C_n$

$C_{_1}, C_{_2}, C_{_3}, \dots, C_{_n}$

Forms An Emotional Trajectory.

XII. EMOTIONAL TRAJECTORY ANALYSIS

Consider a hypothetical character whose emotional profile across six chapters is:

Chapter	Dominant emotional state	Secondary state
1	Hope	Love
2	Joy	Trust

3	Fear	Confusion
4	Anger	Grief
5	Sadness	Guilt
6	Hope	Relief

This trajectory can be interpreted as:

Hope → Joy → Fear → Anger → Grief → Hope

The computational model identifies the pattern; literary interpretation explains why the transformation occurs. This distinction is important. Machine learning can reveal that emotional language changes, but literary scholarship is required to interpret the symbolic, cultural, narrative, and philosophical significance of the change.

XIII. ILLUSTRATIVE EXAMPLE

Consider the following fictional passage:

“Maya stood beside the window long after everyone had left. She remembered her father’s voice and felt an ache that she could not explain. Yet somewhere within her, the possibility of tomorrow remained.”[12] A conventional sentiment classifier might classify the passage as predominantly negative. The proposed framework produces a richer representation:

Character	Evidence	Emotion	Psychological Indicator
Maya	“Long After Everyone Had Left”	—	Isolation
Maya	“Remembered Her Father’s Voice”	Sadness/Grief	Bereavement
Maya	“An Ache”	Sadness	Emotional Pain
Maya	“Possibility of Tomorrow”	Hope	Future Orientation

Thus, the passage contains both **grief and hope**.

A binary sentiment classifier could overlook this coexistence.

XIV. Experimental Evaluation Framework

A genuine empirical implementation should divide annotated data into:

- Training Set: 70%;
- Validation Set: 15%;
- Test Set: 15%.

A character-aware test set should be manually annotated by at least two or three trained annotators.

Each annotation should contain:

1. Character;
2. Emotional Expression;
3. Emotion Category;

4. Intensity;
5. Textual Evidence;
6. Psychological-State Indicator Where Appropriate.

XV. EVALUATION METRICS

The following metrics should be reported.

Accuracy

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

Accuracy is useful but may be misleading when emotion classes are imbalanced.

Precision

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

Recall

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

F1 Score

$$F1 = 2 * \{(\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})\}$$

Where,

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

For multi-class emotion classification, macro-F1 should be reported because it gives equal weight to individual categories.

Additional metrics should include:

- Micro-F1;
- Macro-F1;
- Weighted-F1;
- Confusion Matrix;
- Cohen's Kappa Or Krippendorff's Alpha For Human Agreement.

XVI. PROPOSED COMPARATIVE EXPERIMENT

The following systems should be compared:

Model -A: Lexicon baseline

NRC- Based Emotion Detection.

Model-B: TF-IDF + Logistic Regression

Traditional supervised machine learning.

Model-C: TF-IDF + SVM

A Strong Classical Baseline.

Model-D: BiLSTM

Sequential Deep-Learning Model.

Model-E: BERT

Contextual Transformer Model.

Model-F: Character-aware BERT

BERT plus character identification and co reference information.

Model-G: Hybrid model

NRC lexical features + transformer embeddings + character/context features.

The central empirical question is whether adding character information and narrative context improves emotion attribution.

XVII. RESULTS FRAMEWORK

Because this manuscript is being developed without an accompanying experimental dataset, numerical results should **not be invented**. The final empirical version should contain a table in the following form after model training:

Model	Accuracy	Macro-F1	Micro-F1	Precision	Recall
NRC Lexicon	[Actual]	[Actual]	[Actual]	[Actual]	[Actual]
TF-IDF + LR	[Actual]	[Actual]	[Actual]	[Actual]	[Actual]
TF-IDF + SVM	[Actual]	[Actual]	[Actual]	[Actual]	[Actual]
BiLSTM	[Actual]	[Actual]	[Actual]	[Actual]	[Actual]
BERT	[Actual]	[Actual]	[Actual]	[Actual]	[Actual]
Character-aware BERT	[Actual]	[Actual]	[Actual]	[Actual]	[Actual]
Hybrid model	[Actual]	[Actual]	[Actual]	[Actual]	[Actual]

This design allows the paper to become an empirically complete Scopus submission once the models have been executed.

XVIII. DISCUSSION

The proposed framework highlights several important issues in computational literary studies.

First, literary emotion is fundamentally contextual. A word such as *cold* can describe temperature, personality, social rejection, or emotional detachment. Therefore, lexical matching alone is insufficient. Second, emotion is character-specific. The same sentence can contain emotional states belonging to different characters. Character identification and co reference resolution are consequently central to reliable analysis. Third, emotions can coexist. A character may experience grief and hope simultaneously. Fine-grained multi-label classification is therefore preferable to forcing every passage into a single category. Fourth, literary emotion is frequently indirect.[13] A character may never explicitly say “I am afraid.” Instead, fear can be represented through trembling, silence, avoidance, bodily sensations, fragmented speech, or changes in behavior. Fifth, psychological states develop over time. A character who begins a novel with confidence may gradually become isolated, anxious, angry, and finally reconciled. This development is more meaningful than isolated sentence-level predictions.

XIX. MACHINE LEARNING AND CLOSE READING

The proposed framework should not be understood as a replacement for close reading. Instead:

Close Reading Asks:

“What does this passage mean?”

Machine Learning Asks:

“What patterns occur across this corpus?”

The two approaches can therefore operate together.

For example, a model may identify an unusually high concentration of fear-related passages surrounding a character during the middle third of a novel. A scholar can then return to those passages and examine:

- Narrative Perspective;
- Metaphor;
- Symbolism;
- Dialogue;
- Historical Context;
- Gender Representation;
- Cultural Context;
- Psychological Characterization.

This creates an iterative research process:

Computational Detection→**Close Reading**→**Interpretation**→**Computational Validation**
Computational \ Detection \right arrow Close \ Reading \right arrow Interpretation \right arrow Computational \ Validation

Such an approach can strengthen both quantitative and qualitative literary scholarship.

XX. Explainable AI for Literary Analysis

Explain ability is especially important because literary scholars may question automated classifications.

The proposed system should provide evidence such as:

Character: Maya

Predicted Emotion: Sadness

Confidence: 0.81

Evidence: “remembered her father's voice,” “an ache,” “everyone had left”

Rather than providing only a label, the model should identify textual evidence.

Techniques such as:

- Attention Visualization;
- Shape;
- Lime;
- Token Attribution;
- Integrated Gradients

Can help explain model predictions. Explain ability can therefore transform machine learning from a “black box” into a research assistant for literary interpretation.

XXI. ETHICAL CONSIDERATIONS

The proposed research raises important ethical and interpretive issues.

i. Fiction is not Clinical Evidence

A model detecting sadness-related language in a fictional character does not diagnose depression.

ii. Cultural Variation

Emotional expression varies across cultures, historical periods, genres, and linguistic communities.

iii. Gender and Social Bias

Training datasets may contain social stereotypes that influence emotion predictions.

iv. Historical Language

Words used in nineteenth-century literature may have different emotional meanings from contemporary language.

v. Annotation Subjectivity

Different readers may interpret the same passage differently.

vi. Algorithmic Bias

Research has demonstrated that even widely used emotion lexicons may contain problematic or contextually inappropriate associations. (ACL Anthology)[14]

Consequently, computational predictions should remain open to human correction.

XXII. IMPLICATIONS OF THE STUDY

i. Literary Studies

The framework can help scholars examine character development across large corpora.

ii. Digital Humanities

It provides a computational method for studying affective patterns in literature.

iii. NLP

Literary narratives constitute a challenging domain for contextual emotion detection.

iv. Artificial Intelligence

The study demonstrates how AI can support interpretive disciplines without eliminating human interpretation.

v. Language Education

Emotion-aware literary analysis can potentially support students in understanding characterization, narrative perspective, and textual evidence.

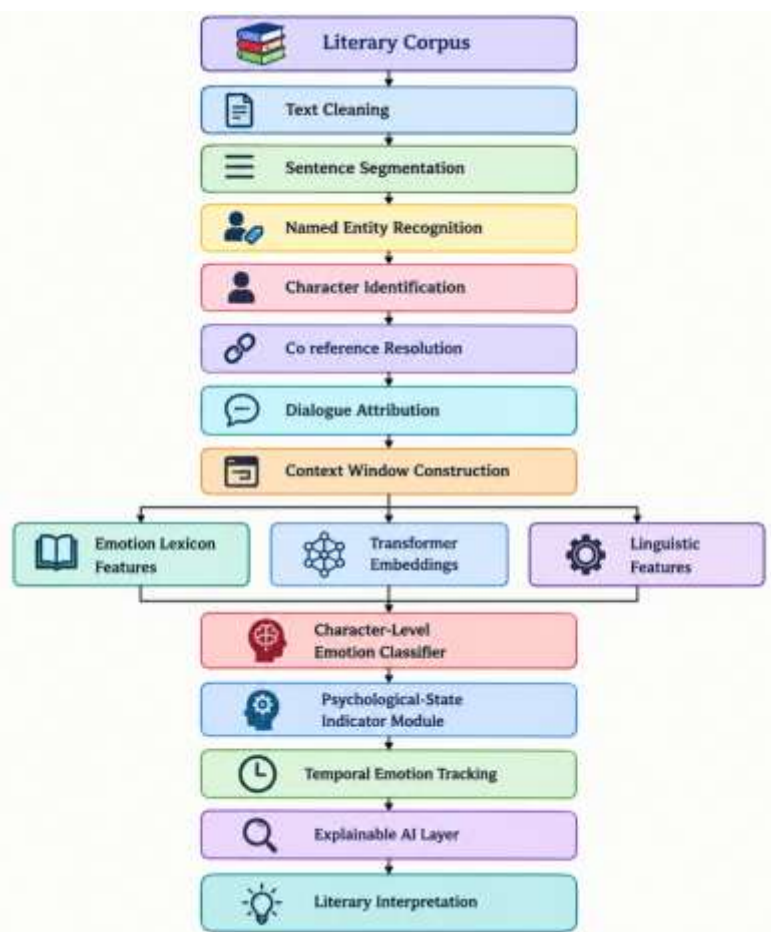
vi. Interdisciplinary Research

The framework connects:

English Literature + NLP + Machine Learning + Psychology + Digital Humanities

XXIII. PROPOSED COMPUTATIONAL ARCHITECTURE

The complete architecture can be represented as:



This architecture provides a bridge between computational prediction and humanistic interpretation.

XXIV. Conclusion

This paper has proposed a machine-learning-based framework for detecting character emotions and psychologically significant states in literary narratives. The study addresses a major challenge in computational literary analysis: emotional meaning cannot be adequately represented through simple positive/negative sentiment classification.[15] A character-aware system must identify the emotional subject, understand contextual meaning, distinguish narration from dialogue, resolve references, and track emotional change across narrative time. The proposed framework therefore integrates character recognition, co reference resolution, lexical emotion resources, traditional machine learning, transformer-based contextual representations, psychological-state indicators, and temporal modeling.[16]

The study also argues for a clear distinction between computational detection of psychological indicators and clinical psychological diagnosis. Literary characters are textual constructs, and computational models should identify patterns in textual representation rather than make medical claims. Existing resources such as the NRC Emotion Lexicon provide interpretable lexical information, while fine-grained datasets such as Go Emotions demonstrate the potential of modern emotion classification. (National Research Council Canada)[17] Research specifically addressing narrative affect further establishes the importance of identifying the character associated with an emotional expression. The principal contribution of the proposed framework is its movement from sentence-level emotion detection toward character-centered emotional trajectory analysis. Such a model can identify patterns such as:

Hope → Love → Fear → Anger → Grief → Reconciliation

The future empirical implementation should compare lexical, traditional machine-learning, neural, transformer, and character-aware transformer models using manually annotated literary data. Performance should be evaluated through macro-F1, micro-F1, precision, recall, accuracy, and human annotation agreement. [18] Importantly, the final empirical paper should report only experimentally obtained values. Ultimately, machine learning should be viewed not as a substitute for literary interpretation but as a methodological instrument capable of revealing large-scale affective patterns. [19] When combined with close reading and humanistic theory, computational emotion analysis can contribute to a richer understanding of character psychology, narrative development, and the representation of human experience in literature.

REFERENCES

- [1] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- [2] Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1–8.
- [3] Boyd-Graber, J., Hu, Y., & Mimno, D. (2017). Applications of topic models. *Foundations and Trends in Information Retrieval*, 11(2–3), 143–296.
- [4] Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., & Ravi, S. (2020). GoEmotions: A dataset of fine-grained emotions. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 4040–4054. (ACL Anthology)
- [5] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT*, 4171–4186.
- [6] Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6(3–4), 169–200.
- [7] Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4), 82–89.
- [8] Jurafsky, D., & Martin, J. H. (2023). *Speech and Language Processing* (3rd ed., draft). Pearson.
- [9] Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool.
- [10] Mohammad, S. M. (2012). From once upon a time to happily ever after: Tracking emotions in novels and fairy tales. *Decision Support Systems*, 53(4), 730–741.
- [11] Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3), 436–465.
- [12] Mohammad, S. M. (2013). NRC Emotion Lexicon. National Research Council Canada. The resource associates English words with eight emotions and positive/negative sentiment.
- [13] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135.
- [14] Plutchik, R. (1980). *Emotion: A Psychoevolutionary Synthesis*. Harper & Row.
- [15] Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing research based on function-component-representation framework. *IEEE Transactions on Affective Computing*, 8(1), 1–15.
- [16] Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. *Proceedings of EMNLP*, 1631–1642.
- [17] Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24–54.
- [18] Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. *Proceedings of ACL*, 417–424.

- [19] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
- [20] Al-Rawi, A., & others. (2023). Artificial intelligence approach for detection and classification of depression among refugees in selected diasporic novels. *Social Sciences & Humanities Open*, 8(1), 100558. The study demonstrates the use of NLP, sentiment analysis, topic modelling, and machine-learning classifiers for psychologically themed literary analysis. (ScienceDirect)
- [21] Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., & Ravi, S. (2020). *GoEmotions: A Dataset of Fine-Grained Emotions*. ACL Anthology. The dataset contains approximately 58,000 English comments annotated across 27 emotion categories plus neutral. (ACL Anthology)