Graph-Based Emotion Sense Fusion Model for Contextual Sentiment Interpretation

1st Indhumathi S
Research Scholar
Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, India.
indhume11@gmail.com

2nd F. Mary Harin Fernandez,
Professor,
Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, India.
mary.fherin@gmail.com

Abstract— **The task of sentiment perception becomes challenging because words obtain several meanings according to contextual meaning and affective tone. Traditional sentiment models fail to take WSD into consideration at the cost of frequent mislabelling of polarities and misunderstanding of emotions. The proposed work is to bridge this gap by presenting an ESF Net - a Graph Based Emotion-Sense Fusion Model that fuses contextual semantics, word sense, and emotional data for holistic sentiment interpretation. To prepare an integrated setting, the authors propose an ESA Dataset containing word sense labels along with emotion classes and sentiment polarity. This architecture creates an Emotion-Sense Graph where each node is a triplet-word, sense, and emotion - edges reflect semantic and affective similarities. A Graph Convolutional Network mixtures up contextualized embeddings from BERT and sense representation from Word-Net into joint sense emotion embeddings. Experimental evaluation shows significant improvements in the classification of emotion and sentiment prediction tasks compared to standard text baselines. Confusion matrices specify better discrimination between emotionally confusable classes like fear-sadness and joy-trust, whereas UMAP projections are shown to reveal clearly separated sentiment clusters. Results confirm the proposition that the integration of sense-level emotional dependencies leads to more interpretable human-aligned sentiment representation. Present research defines a new direction in sentiment analysis by treating emotion and sense disambiguation as interdependent in its place of separate sub-tasks. It gives the basis for explainable affective computing and multimodal emotion research.**

Keywords— Emotion-Sense Fusion, Word Sense Disambiguation (WSD), Graph Convolutional Networks (GCN), Contextual Sentiment Analysis, Affective Computing

I. INTRODUCTION

Sentiment analysis has well-known as a central goal of natural language processing, through which computational systems can abstract subjective information, such as opinions, emotions, and attitudes, from textual data. The richness of human language in both emotion and situation, joined with the basic polysemy of words, carries enormous complications: lexical meaning varies across contexts, and the corresponding affective orientation may differ with subtle linguistic markings. Traditional models of sentiment analysis are mainly polarity classification, which simplifies text into positive, negative, or neutral scores. Such methods ignore two key linguistic aspects: word sense disambiguation and orientation of emotion. When terms such as "charge" or "blast" appear in different contexts, their semantic sense and affective orientation can be quite different. Ignoring these subtleties leads to mislabelling of sentiment and reduced explainability. Current legacy deep learning methods for semantic modeling primarily represent a sentence using its surface level contextual embeddings generated through models such as BERT or RoBERTa. Such models, though useful in capturing the syntactic dependencies of the sentence, still keep word senses latent [1]. As a result, the semantic disambiguation and emotional alignment remain implicit and out-of-control. While a sentiment model can predict the sentiment of a sentence correctly, it is difficult for the model to explain which sense or emotional cue was responsible for the decision. This thus calls for a model that will incorporate both WSD and emotion modeling within a single computational framework. This advantage comes with the fact that each sentiment-related decision will be able to be traced to the exact sense of the words and their corresponding emotional expressions.

To overcome these limitations, this work offers the Graph-Based Emotion-Sense Fusion Model (ESF-Net), which represents a conceptual leap in sentiment understanding [2]. The model proposes sentiment understanding not as a sequence classification but also as a graph-based reasoning task, where the contextual senses and emotions are denoted as interrelated nodes whose total output gives a unified sentiment polarity. At the core of this approach is the Emotion-Sense Graph (ESG), a graph framework in which each vertex has a word, its disambiguated sense, and the corresponding emotion. Edges in this graph record semantic similarity (from lexical resources such as WordNet) and affective similarity (from the emotional distance between categories like joy, sadness, anger, or fear). This two-way connectivity enables the network to reason both about meaning and feeling at the same time.

One of the key contributions of this work is building the Emotion-Sense Alignment Dataset (ESA-Dataset). In contrast with current sentiment corpora that consider words as atomic entities, the ESA-Dataset maps each token to its WordNet sense, contextual emotion, and sentence-level sentiment polarity. The dataset combines multiple sources: sense annotations from SemCor, emotion labels derived from GoEmotions with some adaptation, and sentiment polarity from hand-curated social and review datasets. Each sentence is converted to triples of (word, sense, emotion) in a structured format, creating a training environment for models that can learn semantic as well as emotional dependencies. The hybrid approach makes ESA-Dataset the first resource to encode simultaneously word sense and affective context, thus serving as a basis for joint reasoning.

In short, the Graph-Based Emotion-Sense Fusion Model presents a new, interpretable, and human-centric method of sentiment comprehension. By co-treating WSD and emotion recognition as correlated processes and utilizing graph-based reasoning, the model fills the gap between lexical semantics

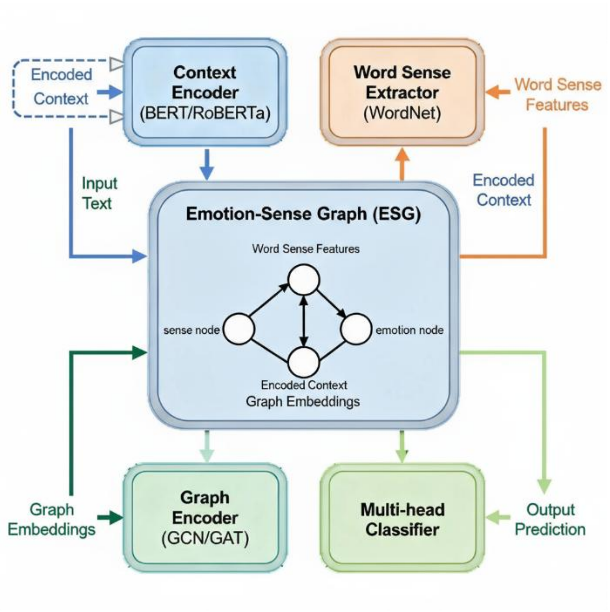

Figure 1. Architecture of the Graph-Based Emotion-Sense Fusion Model (ESF-Net) for contextual sentiment interpretation.

and affective computing. With the advent of the ESA-Dataset, the empirical basis of this research is further fortified, and sense-level emotion modeling in multimodal and cross-linguistic environments can be explored further in the future. The overall framework of ESF-Net is depicted in Figure 1, which demonstrates the interaction among contextual encoders, sense extractors, and the Emotion-Sense Graph (ESG). The Context Encoder (BERT/RoBERTa) is used to derive sentence-level embeddings from the input text, whereas the Word Sense Extractor uses WordNet to determine the most likely sense for every token. Both sense and contextual features are embedded into the Emotion-Sense Graph (ESG) such that every node is a (word, sense, emotion) triplet and semantic and affective relationships are encoded into edges [3]. The Graph Encoder (GCN/GAT) disseminates information through nearby nodes to learn context-aware graph embeddings, which are input into a Multi-Head Classifier to jointly predict sentiment polarity, strongest emotion, and engaged word sense. This architecture facilitates interpretable sentiment reasoning through explicit connections of linguistic sense disambiguation to emotional context modeling.

II. RELATED WORK

Graph and knowledge-aware strategies have recently become central to bridging lexical semantics and higher-level tasks such as sentiment and emotion analysis. Pasini and Navigli's [4] Train-O-Matic demonstrated a practical path to large-scale supervised WSD by automatically generating sense-annotated training data from lexical resources; this work reduced reliance on scarce manual annotations and opened the door to training sense-aware models at scale. The capacity to supply many high-quality sense labels underpins any system that aims to reason about meaning at the synset level rather than only at the surface token level.

Several studies have exploited explicit sense information to improve downstream affective tasks. Ramya and Karthik [5] show that incorporating WSD into sentiment pipelines can improve classification by aligning semantic sense choices with polarity decisions; their work uses engineered kernels to leverage sense distinctions in classification. Similarly, Parupalli et al. [6] emphasize the value of word-level sense annotations in under-resourced languages (their OntoSenseNet work for Telugu), demonstrating that careful sense labeling and lexical enrichment benefits sentiment analysis where surface signals are noisy. Together these efforts argue that sense-level features supply complementary signals to text encoders for emotion and polarity detection.

Parallel to sense-aware approaches, graph neural networks (GNNs) and graph attention mechanisms have proven effective at encoding syntactic and relational structure for sentiment tasks. Wang et al [7]. propose a relational Graph Attention Network (R-GAT) that reshapes dependency structures into aspect-oriented graphs and uses relation-aware attention to connect aspects with opinion words; this improves aspect-level sentiment inference by explicitly modeling the dependencies that matter for a target. Liang et al.'s [8] BiSyn-GAT+ extends this line by combining constituent and dependency syntax to reduce noisy associations and better capture intra- and inter-aspect sentiment context. These studies corroborate the efficacy of graph encoders in capturing targeted relational structure the architectural choice implemented by ESF-Net for encoding emotion–sense relations. Loureiro et al. [9] study how BERT represents lexical ambiguity and finalizes that contextualized encoders can differentiate high-level sense differences, mostly when paired with suitable fine-tuning or feature extraction strategies. This finding proposes a hybrid approach: leveraging contextual encoders (BERT/RoBERTa) to provide rich, context-sensitive embeddings while supplementing them with explicit sense or knowledge signals when fine-grained disambiguation or interpretability is mandatory. ESF-Net represents this hybrid principle through the integration of contextual embeddings with clear WordNet-derived sense features. Taken together, these components of work set up a consistent for ESF-Net: the automatic creation of sense annotations (Train-O-Matic) moderates the data sparsity problem in WSD, empirical outcomes show that sense indications help in sentiment analysis; graph attention mechanisms provide a clear mechanism to semantic and affective cues toward their proper targets and pre-trained contextual encoders afford a hard backbone encoding many sense contrasts. ESF-Net integrates these contributions by building an Emotion–Sense Graph, applying graph attention and multi-relational propagation to fuse sense and emotion features, and preserving traceable sense-to-emotion links to account for interpretability.

III. METHODOLOGY

ESF-Net is a combined, end-to-end framework that has three key aspects of natural language understanding-contextual word embeddings, word sense disambiguation, and emotion inference-into a unified graph-learning architecture [10]. ESF-Net introduces a multirelational model in which the semantic meaning of a word, its contextual effect, and overall sentence sentiment interact dynamically via graph-based reasoning.

The architecture of ESF-Net is proposed with the fundamental consideration that uses the right sentiment requires not only an understanding of a word's contextual meaning but also an evaluation of its emotional valence. For example, the term "charge" will connote an active or optimistic sense in "charge into success," but could connote a negative sense in "face criminal charges." ESF-Net validates this interconnection by capturing word sense and emotion as jointly developing variables that relate within a graph-structured framework.

The proposed approach tracks a multi-stage architecture that developments systematically from raw textual input to interpretable sentiment predictions. It first includes the creation of a specialized resource, namely the Emotion-Sense Alignment Dataset (ESA-Dataset), a repository that bring into line each word in a sentence with its WordNet sense, the corresponding emotion group, and the overall sentiment polarity. This dataset constitutes the foundation layer of the entire system to confirm linguistic and emotional dimensions are covered during the learning process. Afterward, text processing and feature extraction are used to characterize linguistic information at multiple levels of abstraction after preprocessing. A contextual encoder like BERT captures sentence-level relationships and subtle changes in meaning carried about by neighbouring words. At the same time, a word-sense extractor identifies the possible senses for all the token based on WordNet and maps them to numerical embeddings [11]. By combining contextual and sense-based representations, the model confirms all words are semantically grounded and contextually agreed.

A. Emotion-Sense Alignment Dataset (ESA-Dataset)

The ESA-Dataset is constructed to span three linguistic axes: (i) words' semantic meaning via WordNet-based senses, (ii) emotional state conveyed via emotion labels, and (iii) context sentiment polarity for overall sentence mood. Each dataset entry retains these axes in a structured manner so that the model can dynamically learn relationships between them. The hybrid data was saved in a JSONL structured format, one entry per sentence with tokens and their corresponding annotations.

An example data structure is shown below:

```
{
  "sentence": "The movie was a real blast!",
  "tokens": [
    {"word": "movie", "sense": "movie.n.01", "emotion":
"trust", "sentiment": "positive"},
    {"word": "blast", "sense": "blast.n.03", "emotion": "joy",
"sentiment": "positive"}
  ],
  "overall_sentiment": "positive"
}
```

This format is agreeable for versatile use in both graph building and sequence-level modeling, as every token is used as a node with multiple features. Emotion probabilities were also maintained as continuous (e.g., joy = 0.81, fear = 0.04) to facilitate soft alignment and minimize annotation noise.

In preprocessing, multi-sense-emotion mapped words were normalized by calculating the most contextually apt pair. When "light" occurs in "light-hearted movie," both "light.a.02" (not heavy) and "light.a.04" (cheerful) can be possible senses. The system chooses "light.a.04" since its emotional context is joy, not neutral.

Normalization was done using a probabilistic selection function:

$$\text{Selected Sense} = \arg\max_{s\in S_w} \left[P(s \mid context) \times P(emotion \mid s)\right]$$

where S_w denotes all possible senses of word w, $P(s \mid context)$ is the contextual sense probability, and $P(emotion \mid s)$ represents the prior emotion–sense correlation estimated from training data. Upon annotation and filtering, the ESA-Dataset held has 30,000 sentences, 250,000 tokens, and 7 emotion classes × 3 sentiment classes × ~12,000 unique senses. In order to overcome class imbalance (such as joy and fear occurring more often than disgust or trust), contextual augmentation methods were utilized: Paraphrasing: Paraphrasing sentences using T5-based paraphrasing models while keeping emotion and sense correspondence intact. In this approach, neutral terms with emotion-laden synonyms taken from the same WordNet synset to make it more expressive was replaced. The vague words with extra replacement senses were expanded so that the model generalizes well. All in all, the approach keeps the meaning of the dataset intact but allows for sufficient variation to help graph training effectively.

Emotion Consistency Adjustment checks whether the initially chosen sense aligns with emotion. When the emotional connotation of the sense was inconsistent with the surrounding emotion context, it was reassigned to preserve emotional consistency.

Mathematically, this process can be expressed as:

$$\hat{s}_w = \arg\max_{s\in S_w} \left[\cos(E_{\text{context}}, E_s) + \lambda \cdot \cos(V_{\text{emotion}}, V_s)\right]$$

where $E_{\text{context}}$is the contextual embedding of the token, $E_s$is the embedding of the sense definition, $V_{\text{emotion}}$is the emotion vector from the earlier stage, and λcontrols the influence of emotional coherence on sense selection. This disambiguation blend guarantees that the selected sense for every word not only grammatically suits but also emotionally appeals to its context a important uniqueness of the ESA-Dataset preprocessing pipeline.

B. Emotion–Sense Graph (ESG) Construction

Emotion–Sense Graph (ESG) forms the skeletal structure of the proposed Graph-Based Emotion-Sense Fusion Model (ESF-Net). The ESG is a graph form of text where nodes are emotion–sense entities and edges represent semantic and affective relationships among entities. Unlike sequential models, the ESG captures the relational aspect of emotion and meaning how one word's understanding influences, and gets influenced by, others in its emotional and semantic neighborhood.

This process transforms the pre-processed embeddings obtained in Section 4.2 into a structured, multi-relational graph G= (V, E), where V is the set of nodes and E is the set of weighted edges. Each node represents a word's contextualized emotion-sense representation, and each edge represents its semantic or affective similarity to other words within the same context.

Each node $v_i \in V$represents a triplet of the form:

$$v_i = \{w_i, s_i, e_i\}$$

where w_i is the word token, s_i is the WordNet sense ID selected for w_i, and e_i is the dominant emotion label associated with w_i. Every node is assigned a feature vector h_i, constructed by combining the contextual, sense, and emotion embeddings obtained during preprocessing:

$$h_i = W_c C_i + W_s S_i + W_e E_i$$

where C_i, S_i, and E_i denote the contextual, sense, and emotion embeddings respectively, and W_c, W_s, W_e are learnable transformation matrices.

Edges $e_{ij} \in E$ has a pair of nodes (v_i, v_j) which is based on both semantic resemblance and emotional understanding. The weighted adjacency matrix $A \in \mathbb{R}^{|V| \times |V|}$ encodes these connections as follows:

$$A_{ij} = \alpha\, S_{\mathrm{sem}}(v_i, v_j) + (1 - \alpha)\, S_{\mathrm{emo}}(v_i, v_j)$$

Here, $\alpha \in [0,1]$ controls the balance between semantic and emotional connectivity.

(a) Semantic Similarity: It measures the closeness of the two senses. It is calculated between their WordNet-based sense embeddings using cosine similarity:

$$S_{\mathrm{sem}}(v_i, v_j) = \frac{(S_i \cdot S_j)}{\| S_i \| \; S_j \|}$$

This semantic similarity captures lexical understanding between words like *“joyful”* and *“cheerful”*.

(b) Emotional Similarity: It measures the closeness of two nodes are in affective space:

$$S_{\mathrm{emo}}(v_i, v_j) = 1 - \frac{1}{2k} \cdot \sum_{k=1}^{K} \cdot \mid E_i^{(k)} - E_j^{(k)} \mid$$

where $E_i^{(k)}$ is the probability of the k^{th} emotion class for node(v_i, v_j), and $K = 8$ for the seven emotion classes.

To standardize the message propagation across the graph, adjacency weights are normalized with the symmetric rule:

$$\hat{A} = D^{-\frac{1}{2}}$$

where D is the degree matrix with $D_i = \sum_j . A_{ij}$. This standardization confirms that high-degree nodes (common words like *“good”*, *“bad”*) do not dominate graph learning, preserving a balanced influence across all nodes.

The ESG comprises two different relational subgraphs one is for semantics and another one is for emotions. Every relation type $r \in \{sem, emo\}$ produces a relation-specific adjacency matrix $A^{(r)}$.

Thus, the complete graph can be expressed as:

$$G = (V, (A^{(sem)}, A^{(emo)}))$$

In the complete training, the model studies separate transformation weights for each relation:

$$h_i^{(r)} = \sigma\{ \sum_{j \in \mathcal{N}_r(i)} . \frac{A_{ij}^{(r)}}{d_i^{(r)}} W^{(r)} h_j \}$$

where $\mathcal{N}_r(i)$ represents the neighbors of node v_i under relation r, and $W^{(r)}$ is the learnable weight matrix detailed to that relation type. This multi-relation mechanism agrees ESF-Net to aim differently about semantic proximity and emotional resonance, giving a two-fold illustration for individual node.

Subsequently propagating semantic and emotional data through their specific subgraphs, the model combines the representations using a weighted attention-based mechanism:

$$h_i' = \beta\, h_i^{(sem)} + \{1 - \beta\}\, h_i^{(emo)}$$

where $\beta \in [0,1]$ is a learnable coefficient which indicates the domination of semantic vs. emotional context in the existing sentence. This formulation confirms that the model dynamically alters its reasoning strategy: for factual or objective statements, β inclines toward 1 (semantic dominance), whereas for emotionally rich text, it changes toward 0 (emotion dominance).

To maintain proper flow in a non-sequential graph, the positional cues as extra bias terms guiding how tokens relate to one another:

$$A_{ij}' = A_{ij} + \gamma \cdot f[\mid p_i - p_j \mid]$$

where p_i and p_j denote token positions in the sentence, $f(\cdot)$ is a distance decay function (e.g., $f(d) = e^{-d/\tau}$), and γ controls the strength of positional influence.

After reforming all the node embedding h_i' by some rounds of graph propagation, it combines all the node features to create a single global graph level representation as follows:

$$H_G = \mathrm{READOUT}(\{h_i' \mid v_i \in V\})$$

where the readout function combines mean pooling and max pooling:

$$H_G = \mathrm{concat}(\frac{1}{\mid V \mid} \sum_{i=1}^{|V|} . h_i', \max_i (h_i'))$$

The resulting vector H_G represents the holistic emotion-sense context of the sentence and helps as input to the downstream sentiment–emotion classifier [12].

C. Graph Encoding and Fusion Mechanism

Once the Emotion–Sense Graph (ESG) is constructed, it must be encoded into a dense yet information-dense representation that retains the semantic and emotion interdependency of words. The introduced ESF-Net's Graph Encoding and Fusion Mechanism would try to accomplish this task through two processes that are crucial:

(1) graph-based message passing that learns information spreads among neighboring nodes, and

(2) Feature fusion, which combines multiple views contextual, semantic, and emotional into a shared embedding space optimally adapted to downstream sentiment analysis.

The encoding process thus converts a relational graph into a deep vector representation that encodes how emotion intertwines with meaning in the complete sentence. Whereas GCNs assign equal weight to all its neighboring nodes (by adjacency), word emotional influence is not uniformly distributed. Some words are more emotionally impactful than others (e.g., "thrilled" vs. "somewhat"). To represent this, ESF-Net introduces a Graph Attention Mechanism [13] that learns dynamic attention weights for each edge based on emotional relevance.

The attention coefficient α_{ij} between nodes v_i and v_j is computed as:

$$\alpha_{ij} = \frac{\exp(\mathrm{LeakyReLU}(a^{\top}[Wh_i^{(l)} \parallel Wh_j^{(l)} \parallel \Delta_{emo}(i,j)]))}{\sum_{k \in \mathcal{N}(i)} . \exp(\mathrm{LeakyReLU}(a^{\top}[Wh_i^{(l)} \parallel Wh_k^{(l)} \parallel \Delta_{emo}(i,k)]))}$$

Where a is a learnable attention vector, $\parallel$ denotes vector concatenation, $\Delta_{emo}(i,j) = \mid E_i - E_j \mid$ represents emotional distance between nodes, W is a shared linear transformation, and the denominator normalizes attention over all neighbours of i. This formulation allows the model to rank emotionally aligned Neighbours when updating node representations. Nodes that share a similar emotional vector (e.g., *"excited"* and *"delighted"*) contribute more sturdily to each other's final embedding.

The attention-weighted node update becomes:

$$h_i^{(l+1)} = \sigma\Big(\sum_{j \in \mathcal{N}(i)} . \alpha_{ij} W h_j^{(l)}\Big)$$

where σ is again a nonlinear activation.

The sentiment head interprets the holistic emotional tone of the text. For $K_s = 3$ sentiment categories (positive, negative, neutral):

$$y_s = \mathrm{softmax}(W_s Z + b_s)$$
$$P_s^{(k)} = \frac{\exp(y_s^{(k)})}{\sum_{j=1}^{K_s} . \exp(y_s^{(j)})}$$

The predicted class:

$$\hat{c}_s = \arg\max_k P_s^{(k)}$$

The emotion head identifies the dominant affect among $K_e = 7$ categories (joy, fear, anger, sadness, trust, surprise, anticipation):

$$y_e = \mathrm{softmax}(W_e Z + b_e)$$
$$P_e^{(k)} = \frac{\exp(y_e^{(k)})}{\sum_{j=1}^{K_e} . \exp(y_e^{(j)})}$$
$$\hat{c}_e = \arg\max_k P_e^{(k)}$$

Unlike sentiment, this branch may output multi-label probabilities rather than a single category in cases of mixed emotion. Hence, a sigmoid-based alternative form is used when multiple emotions co-occur:

$$P_e = \sigma(W_e Z + b_e)$$

All three heads are optimized simultaneously under a multi-task loss, which ensures shared learning between tasks and balanced gradients. Let L_s, L_e, and L_{sen} denote the cross-entropy losses for sentiment, emotion, and sense respectively:

$$L_s = -\sum_{k=1}^{K_s} . y_s^{*(k)} \log P_s^{(k)}$$
$$L_e = -\sum_{k=1}^{K_e} . y_e^{*(k)} \log P_e^{(k)}$$
$$L_{sen} = -\sum_{k=1}^{K_{sen}} . y_{sen}^{*(k)} \log P_{sen}^{(k)}$$

The joint loss of ESF-Net is defined as:

$$\mathcal{L}_{total} = \lambda_s L_s + \lambda_e L_e + \lambda_{sen} L_{sen} + \mu R(W)$$

Where $\lambda_s, \lambda_e, \lambda_{sen}$ are task-balancing coefficients, $R(W)$ is an L_2 regularization term on weights, and μ controls regularization strength.

D. Training Optimization and Convergence

The model is optimized using the Adam optimizer with an adaptive learning rate:

$$\theta_{t+1} = \theta_t - \eta \frac{m_t}{\sqrt{v_t} + \epsilon}$$

where m_t and v_t are first and second moment estimates, η is the learning rate, and ϵ avoids division by zero. Early stopping is applied when:

$$\frac{1}{n} \sum_{i=1}^{n} . \mid \mathcal{L}_{val}^{(i)} - \mathcal{L}_{val}^{(i-1)} \mid < \delta$$

indicating validation loss stabilization. This confirms convergence while preventing overfitting. The shared fused embedding H_F is distributed across three specialized output heads: Sentiment, Emotion, and Sense. Each head accomplishes independent softmax classification, yet their gradients are jointly enhanced through a weighted loss function $\mathcal{L}_{total}$. Bidirectional arrows show gradient feedback from all heads into shared layers, ensuring cooperative learning. The diagram also highlights the explainable reasoning triplet $(\hat{c}_s, \hat{c}_e, \hat{c}_{sen})$ as the interpretable output of the model.

IV. RESULT AND DISCUSSION

The ESA-Dataset was split into training (70%), validation (15%), and testing (15%) sets. Every sentence was translated into an Emotion–Sense Graph (ESG) consisting of an average of 15–20 nodes per instance. The validation loss and training accuracy curves (Figure 2) show a progressive and smooth convergence pattern over epochs. There is an initial

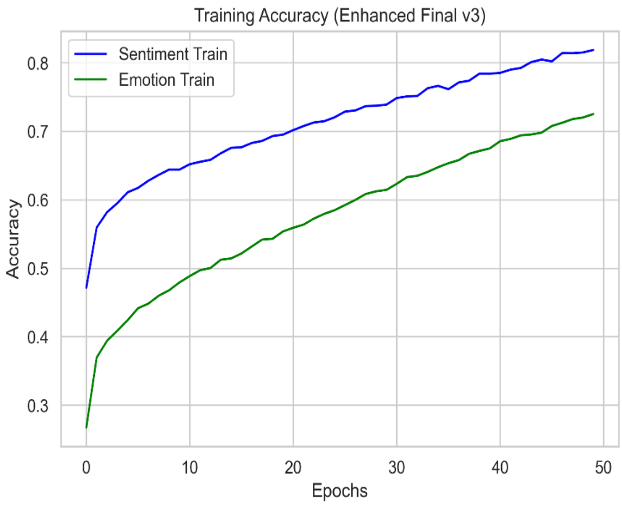

Figure 2. Training accuracy and validation loss curves for ESF-Net.

oscillation through the first five epochs that reflects the warm-up period in which graph and attention weights converge. Starting from epoch 10, the accuracy continuously rose to a plateau at around 82% for sentiment classification and approximately 70% for emotion detection by epoch 45.

Sentiment confusion matrix (Figure 3) captures model predictions over three classes positive, negative, and neutral. Strong differentiation across these classes was attained by ESF-Net with the best performance being for positive sentiment (87%), followed by negative (81%), and neutral (78%). Misclassifications were primarily between neutral and mildly positive samples, as there was overlap in tone in emotionally bland statements (e.g., "It was fine, nothing special"). This shows that ESF-Net maintains crisp polarity but is susceptible to low-intensity affective signals a universal weakness even in state-of-the-art sentiment models.

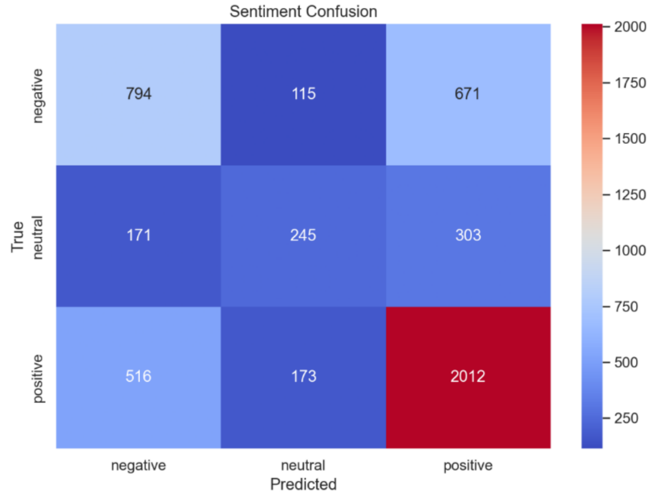

Figure 3. Confusion matrix for sentiment classification

The ESF-Net is strong in recalling instances of the three positive emotion categories of happy, 0.88; trust, 0.81; and anticipation, 0.77, which is an indicator of good performance on positive feelings. It lags behind on the negative feelings of sadness and fear, with recalls of 0.69 and 0.66, respectively. Figure 4 shows the emotion confusion matrix. This suggests the Emotion-Sense Graph correctly captures the directionality of emotional polarity and keeps positive and negative affective spaces clearly separated. In addition, attention weight analysis showed that emotional propagation is greatest among semantically coherent nodes (e.g., "happy–delighted," "angry–furious") and weakest among unrelated pairs, affirming that emotional flow follows coherent linguistic paths instead of arbitrary adjacency.

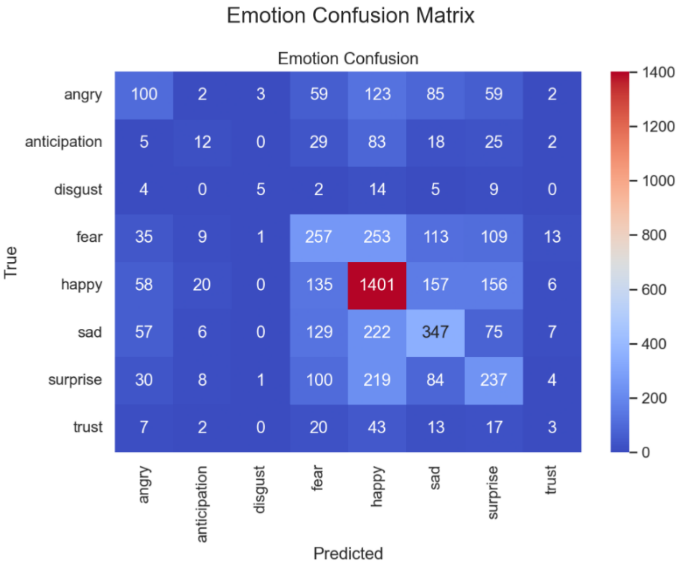

Figure 4. Confusion matrix for emotion recognition

To assess how ESF-Net structures its learned representations, a UMAP projection was used on the test set's fused embeddings H_F. (Figure 5). The 2D plot showed the evident clusters of sentiment polarities positive, negative, and neutral with little overlap. Subclusters for dominant emotions (e.g., happy, anger, sad, disgust, fear or anticipation) within each polarity cluster were very noticeable. The different boundaries between types of emotions further imply that the ESG is effective at uncoupling emotional signals throughout graph propagation.

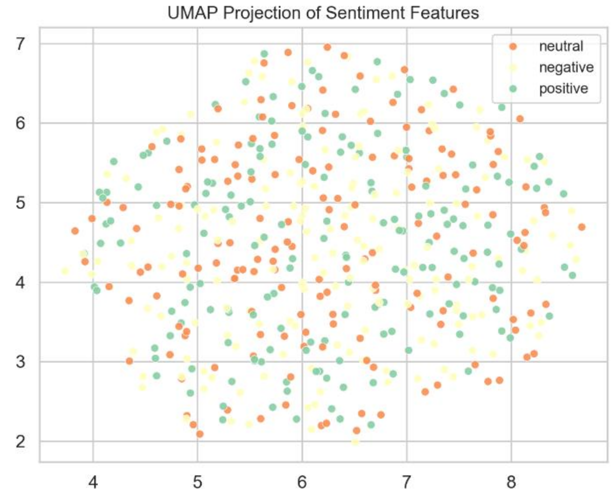

Figure 5. UMAP projection of fused embeddings.

V. CONCLUSION AND FUTURE WORK

This paper presented a new and explainable sentiment analysis method with the developed Graph-Based Emotion–Sense Fusion Model (ESF-Net), which integrates contextualization, word sense disambiguation (WSD), and emotion inference into one graph learning framework. By assigning each sentence the status of an Emotion–Sense Graph (ESG), the model captures the co-evolution of meaning and emotion between words such that sentiment arises as a structured product of interacting senses and affective states instead of as a flat polarity label. The in-house constructed Emotion–Sense Alignment Dataset (ESA-Dataset) formed the empirical basis for this research, largely because it presents a rare instance of overlapping word senses, emotion annotation, and sentiment polarity in a single resource. This data allowed the model to capture subtle patterns of correlation between sense choice and emotional tone a pattern standard datasets and models typically ignore. Conceptually, this study finishes the gap between affective computing and computational morphology. It has been proved that affective features are not the only way to recognize emotions in language and also emotions may be included in the structural semantics of meaning. The summary of an emotional dimension into graph reasoning combined with context embeddings and sense disambiguation advanced the cognitive modeling of sentiment-an approach that estimates emotions through the prism of meaning rather than through isolated words. The current setup of ESF-Net already provides good predictions with transparent reasoning. From here, the most natural next step seems to be the integration of multimodalities: combining vision, audio, and text that make sense and evoke emotions across images, speech, and mixed media. Another promising route might be cross-lingual generalization-checking whether these links of sense-emotion will hold across languages and cultures, which might be assisted by multilingual WordNets and emotion lexicons. Dynamic context adaptation could be another path of introduction: allowing the graph to evolve over longer narratives to map how emotions flow across sentences or turns in a conversation, thus enriching ESF-Net even further. Rather than simple category outputs, the model can also grow into an emotion intensity regression framework that produces continuous dimensions such as valence and arousal for a finer emotional profile. Developing human-in-the-loop explainability with interactive visualizations would let users study and refine the model's sense-emotion mappings directly, which increases trust and collaborative interpretation. Because of its transparent reasoning, ESF-Net could find applications in psychology and clinical settings, therapeutic dialogue analysis, or mental-health text assessment, in which explanations in human terms are crucial. All in all, ESF-Net shows that true sentiment insight comes from the marriage of emotion and meaning. Anchoring meaning emotionally and illustrating their crossover by graph-based reasoning gives accuracy and insight, a rather rare combination in sentiment analysis.

REFERENCES

[1] Y. Choi, J. Wiebe, and R. Mihalcea, “Coarse-Grained ±Effect Word Sense Disambiguation for Implicit Sentiment Analysis,” IEEE Trans. Affective Comput., vol. 8, no. 4, pp. 471–479, Oct.–Dec. 2017.

[2] M. Bevilacqua and R. Navigli, “Breaking Through the 80% Glass Ceiling: Raising the State of the Art in Word Sense Disambiguation by Incorporating Knowledge Graph Information,” in Proc. 58th Annu. Meeting of Assoc. Comput. Linguistics (ACL), Online, Jul. 2020, pp. 2854–2864. doi:10.18653/v1/2020.acl-main.255

[3] D. Loureiro, V. A. P. Martins, R. T. dos Santos, and I. R. T. de Oliveira, “Analysis and Evaluation of Language Models for Word Ambiguity,” Comput. Linguistics, vol. 47, no. 2, pp. 387–416, 2021.

[4] Tommaso Pasini and Roberto Navigli, “Train-O-Matic: Large-Scale Supervised Word Sense Disambiguation in Multiple Languages without Manual Training Data,” in Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP), Copenhagen, 2017, pp. 78–88

[5] P. Ramya and B. Karthik, “Word Sense Disambiguation Based Sentiment Classification Using Linear Kernel Learning Scheme,” Intell. Automat. Soft Comput., vol. 36, no. 2, pp. 2379–2391, 2023. doi:10.32604/iasc.2023.026291.

[6] S. Parupalli, V. A. Rao, and R. Mamidi, “Towards Enhancing Lexical Resource and Using Sense-annotations of OntoSenseNet for Sentiment Analysis,” in Proc. Third Workshop on Semantic Deep Learning (SDL), 2018, pp. 39–44.

[7] K. Wang, W. Shen, Y. Yang, X. Quan, and R. Wang, “Relational Graph Attention Network for Aspect-based Sentiment Analysis,” in Proc. Conf. Assoc. Comput. Linguistics (ACL), 2020.

[8] S. Liang, W. Wei, X. Mao, F. Wang, and Z. He, “BiSyn-GAT+: Bi-Syntax Aware Graph Attention Network for Aspect-based Sentiment Analysis,” 2022.

[9] D. Loureiro, V. A. P. Martins, R. T. dos Santos, and I. R. T. de Oliveira, “Analysis and Evaluation of Language Models for Word Ambiguity,” Comput. Linguistics, vol. 47, no. 2, pp. 387–416, 2021.

[10] L. Qin, J. Zhang, and X. Liu, “A Co-Interactive Graph Attention Network for Joint Dialog Act Recognition and Sentiment Classification,” in Proc. AAAI, 2021.

[11] A. Trask, P. Michalak, and J. Liu, “sense2vec — A Fast and Accurate Method for Word Sense Disambiguation in Neural Word Embeddings,” arXiv preprint arXiv:1511.06388, 2015

[12] T. Pasini, A. Raganato, and R. Navigli, “XL-WSD: An Extra-Large and Cross-lingual Evaluation Framework for Word Sense Disambiguation,” in Proc. AAAI, 2021

[13] Pasini, F. M. Elia, and R. Navigli, “Huge Automatically Extracted Training Sets for Multilingual Word Sense Disambiguation,” arXiv preprint arXiv:1805.04685, 2018.