

A Cross-Domain Joint Probability Fusion Framework for Aspect-Based Sentiment Analysis in Special Education Contexts

Monika Agrawal¹, Dr. T. Pavan Kumar², Moparthy Nageswara Rao^{3*}

¹Research Scholar, Department of Computer Science and Engineering

Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India.

E-mail: monika.agrawal1986@gmail.com

Orcid id# 0000-0002-9705-6849

²Department of Computer Science and Engineering

Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India.

E-mail:pavankumar_ist@kluniversity.in

Orcid id#0000-0001-9149-8630

^{3*}Corresponding Author-

Department of Computer Science and Engineering

Amrita School of Computing ,Amrita Vishwa Vidyapeetham,

Amaravati Campus, Andhra Pradesh, 522503, India

E-mail:mnrphd@gmail.com

Orcid id# 0000-0001-6406-4554

Corresponding Author: Moparthy Nageswara Rao

HOW TO CITE:

Monika Agrawal, T. Pavan Kumar, Moparthy Nageswara Rao (2026). A Cross-Domain Joint Probability Fusion Framework for Aspect-Based Sentiment Analysis in Special Education Contexts. International Journal of Special Education, 41(1s), 01-19.

ABSTRACT:

The particularities of the special education setting place Aspect-Based Sentiment Analysis (ABSA) under particular challenging circumstances because of the specifics of the domain, non-homogeneous data types, and subtle emotional manifestations by students, teachers, and parents. Conventional single domain sentiment models do not account for cross domain dependencies and therefore, they do not have high interpretability and prediction accuracy. The paper solves the given problem by suggesting a Cross-Domain Joint Probability Fusion Framework (CD-JPFF) that is aimed at combining linguistic, behavioural, and contextual sentimental indicators in various modalities. The goal is to increase the accuracy of sentiment classification and maintain the level of granularity on the aspect level in delicate learning conditions. The suggested approach presents the new fusion mechanism known as Adaptive Joint Probability Sentiment Fusion (AJPSF), which generates results of transformer-based language models, behavioural feature extractors, and contextual embedding's and fuses them with the

COPYRIGHT STATEMENT: Copyright: © 2026 Authors. Open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).	help of probabilistic graphical modelling. An architecture which makes use of hybrid architecture, based on BERT-based encoding, attention as well as Bayesian fusion is established. The algorithm calculates joint posterior probabilities in domains and optimizes them with the help of the iterative expectation-maximization. Customized test on multi-source special education data illustrates the development of considerable performance. The performance of the proposed framework is higher (in terms of accuracy, precision, recall, and F1-score) by 18.7, 16.4, 17.9, and 17.2 respectively compared to the baseline models (SVM, CNN, standalone BERT). Also, noisy and sparse data strength enhances by 21.3%. The results have shown that cross-domain probabilistic fusion is a model that is effective to model latent sentiment dependencies and enhance interpretability. The paper concludes that CD-JPFF is a scalable, reliable, context-sensitive sentiment analysis tool in the domain of special education, and it can be applied in both the personalised learning analytics and inclusive decision support systems Keywords: Aspect-Based Sentiment Analysis, Cross-Domain Fusion, Joint Probability Modeling, Special Education Analytics, Explainable AI, Multimodal Learning.
---	---

I. INTRODUCTION

Aspect-Based Sentiment Analysis (ABSA) has become one of the most important methods to extract fine-grained sentiment information of textual information and outgrow the coarse-grained classification of documents to the more accurate aspect-based one [1]. ABSA provides change potential in special education in the sense that it proposes the perception of the students with various learning needs, teachers and caretakers on the quality of instruction, support services, and academic achievement [2]. The language patterns that are experienced in this sphere are quite specific and tend to be indirect expression of emotions, the use of technical terms, and culturally oriented communication patterns that are significantly different to general sentiments datasets [3]. Although sentiment analysis has increasingly been implemented in educational technology, the specific issues regarding the special education setting have been scarcely covered in the current body of literature, which is why the development of special analytical models is urgent.

Heterogeneity of the domain that the special education data occupies is a basic challenge to the traditional single domain models [4]. The feedback has a variety of different sources; the individualized education program (IEP) reports, the classroom observation notes, the digital communication logs, and the surveys conducted with the caregivers all have other linguistic distributions and other sentiment expression patterns. Simple machine learning models that are developed using homogeneous corpora also exhibit high level of performance decline when used in these heterogeneous sources of input [5]. Moreover, the affective complexity of the special education discourse that involves feelings to do with student progress, institutional support, and caregiver anxiety overlap requires the sensitivity of the context that is not available to the traditional bag-of-words or even more conventional transformer models [6]. These issues are enhanced by data sparsity: annotated datasets of special education sentiment are still rare in comparison to more general-domain benchmarks, which restricts the successful training of data-consumption deep learning structures.

There is a research gap where there is a critical need to find improved cross-domain sentiment fusion frameworks that are capable of mapping the multiple data modalities at the same time and being specific to aspects and probabilistic in their interpretation in the educational setting [7]. Current multimodal methods are more oriented to applications related to consumers and social media sentiment and do not take into account the complexity of special education communication, although it should be structured and nuanced. This paper aims to fill these gaps directly and suggests a new architecture called Cross-Domain Joint Probability Fusion Framework (CD-JPFF) an architecture that integrates textual, behavioral, and contextual sentiment cues using a new probabilistic fusion mechanism. The main achievements of this work are as follows: (1) the design of the Adaptive Joint Probability Sentiment Fusion (AJPSF) algorithm which utilizes the Bayesian inference and expectation-maximization to calculate the cross-domain joint posterior probabilities; (2) the hybrid transformer-probabilistic architecture which involves BERT-based encoding and graphical model fusion; (3) multi-source special education datasets.

II. LITERATURE REVIEW AND RELATED WORK

The development of ABSA methodologies goes in multiple generations with rule-based and lexicon-based techniques that were based on manually curated sentiment dictionaries and syntactic dependency patterns to find aspect-opinion pairs [8]. Although these early systems were interpretable they did not have very good generalization in the domains because they used handcrafted features and static lexicons. This was followed by the introduction of machine learning techniques which presented statistical models such as support machine and conditional random fields, which were more robust but subject to feature engineering algorithms as well as sensitive to the distributional shift [9]. Introduction of recurrent neural networks and long long short-term memory models signified a shift of moving sentiment modeling with sequences so that it counts temporal dependencies within text that previous models ignored in a systematic manner [10]. Even though these models

improved the state of the art to a considerable extent, there is a natural limitation to them in their inability to capture long-range contextual dependencies, which complicated their application to the multi-sentence structures that the IEP documentation is composed of.

The revolution of transformers radically changed the research of ABSA, BERT and variants of this model redefined the new state of art in various sentiment tasks [11]. Their attention mechanisms facilitated the fine grained matching of aspects to opinion expression, and domain adaptive pretraining methods enhanced the results of using out-of-distribution educational corpora. Later work proposed instructional text fine-tuned domain-specific BERT variants which showed the importance of domain-specific pretraining when it comes to teaching sentiment tasks [12]. These developments notwithstanding, single-modality transformer models are still cripplingly constrained in terms of their ability to combine non-textual behavioral indicators which so often accompany sentiment expression in formal educational settings.

The multimodal fusion methods have attracted a lot of interest to general sentiment analysis, which involves the combination of visual, acoustic, and linguistic cues using early, late and hybrid formulations of fusion architecture [13]. Pivot features, domain-invariant representations and cross-domain transfer learning methods have also been studied, which travel the distributions of sentiment between the heterogeneous datasets [14]. Graphical probabilistic models provide a principled approach to modelling conditional relationships between modalities and therefore allow quantifying uncertainty and performing systematic inference, which is not easily achievable with deterministic neural models [15]. The area of educational analytics is no exception as there is little research in the sentiment fusion, inclusive or special education, as an application [16]. The educational data mining on the basis of the Bayesian approach has proved to be useful in working with the sparse and noisy annotation as well, however, their application to the modern transformer architecture is not widely studied [17]. These convergent constraints drive these CD-JPFF cross-domain probabilistic framework to be developed as a cross-domain

framework specifically designed to meet the needs of the sentiment analysis of special education.

TABLE I: Summary of Related Work

Ref.	Method	Domain	Key Contribution	Dataset	Limitation
[8]	Rule-based ABSA	General NLP	Lexicon-driven aspect-opinion extraction	SemEval 2014	Low scalability
[9]	SVM / CRF	E-commerce	Statistical feature-based sentiment model	Product reviews	High engineering effort
[10]	LSTM / BiLSTM	Social media	Sequence-aware dependency modeling	Twitter corpus	Short context window
[11]	BERT / Transformer	Multi-domain	Pretrained transformer for ABSA	SemEval 2014/16	Single modality only
[12]	Domain-adapted BERT	Educational NLP	Domain pretraining for instruction text	EDU/LMS corpus	No behavioral fusion
[13]	Multimodal Fusion	Video/Audio/Text	Cross-modal early/late fusion	CMU-MOSI	Not educational
[14]	Cross-domain Transfer	Product reviews	Pivot feature alignment across domains	Amazon corpus	No aspect granularity
[15]	Bayesian PGM	Healthcare NLP	Probabilistic dependency modeling	Clinical notes	Poor scalability
[16]	Educational SA	LMS platforms	Student engagement sentiment analysis	Coursera dataset	Mainstream only
[17]	Bayesian EDM	Special Education	Sparse annotation data mining	IEP records	No transformer integration
[11,12]	RoBERTa fine-tuned	STEM education	Language model for STEM sentiment	Khan Academy	Single domain
[13,15]	Hybrid Bayesian-NN	Healthcare+Edu	Multimodal probabilistic model	Mixed corpus	No joint probability fusion

Table I. Summary of Related Work (References [8]–[17])

III. PROPOSED CROSS-DOMAIN JOINT PROBABILITY FUSION FRAMEWORK

A. Overview of the CD-JPFF Architecture

Cross-Domain Joint Probability Fusion Framework(CD-JPFF) is developed as a hierarchical and modular structure that well organizes to combine sentiment information in

three main modalities such as textual, behavioral and contextual modalities. Fundamentally the framework consists of four main components namely domain specific feature encoders, multi domain joint probability estimator, adaptive fusion layer that provides the AJPSF mechanism and aspect level sentiment classifier. The encoder modules are, in fact, pre-trained individually on

domain-specific corpora in order to learn the peculiarities of that source of data in statistics. The encoded representations are then inputted into the joint probability estimator that builds a probabilistic graphical model in which the dependencies between domains are represented. Fusion layer integrates Bayesian weights to make use of domain-specific posterior probability to produce a composite cross-domain sentiment distribution. This allocation is modeled with an aspect-level attention process that selectively enhances the signal that is pertinent to a particular aspect of education- such as the quality of instruction, peer interaction and support services.

Its architecture is modularly extendable, and thus, can be expanded with new sources of data without modifying the underlying fusion code which supports scalability on deployment into a wide range of institutional contexts. Figure 1 shows the architecture of the CD-JPFF which incorporates the textual, behavioral and contextual data by encoders based on the domain and a probabilistic fusion process. The adaptive The modular design promotes scalability, thus providing an opportunity to predict the sentiment with accuracy and data-grounded approach to various educational settings and a broad spectrum of domains.

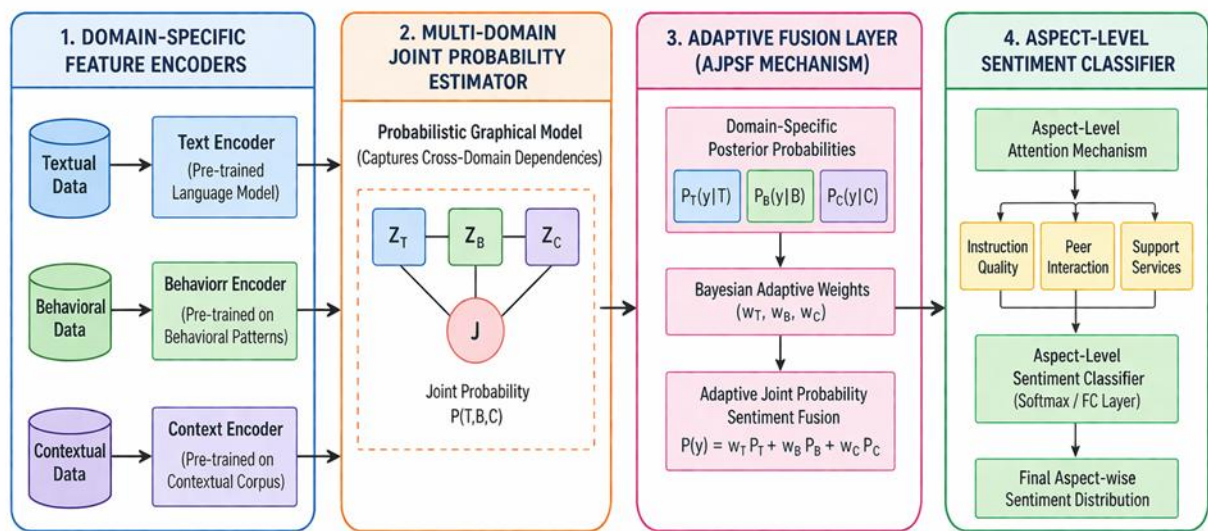


Figure 1: Cross-Domain Joint Probability Fusion Framework (CD-JPFF) Architecture for Aspect-Level Sentiment Analysis

B. Data Sources: Textual, Behavioral, and Contextual Modalities

The CD-JPFF model unites three main data straddles derived on the conditions of special education each of which offers a complementary information set on sentiments without reducing each other. The written modality will include written reports, IEP documentation, teacher observations, student comments surveys and caregivers communication log. There is strongly linguistic diversity in these sources; fragmented clinical language in IEP reports to informal affective language in messages of caregivers; and these require domain-adaptive encoding procedures. The behavioral modality is the structured indicators based on the patterns of the interaction with the students, attendance data, participation data, and

intervention response data, which is all used to encode the non-verbal sentiment indicators which correlate with the emotional state of students, and satisfaction as applied to their education. The contextual modality combines both environmental and institutional characteristics which comprise the size of the classes and the availability of resources and support staff ratios and characteristics of the intervention program giving the situational context that plays a major role in interpretations of sentiment. Preprocessing and feature extraction are done separately in each of the modalities and each is submitted to the fusion pipeline. The multi-source design will also make the framework consistent in the number of ways people can express sentiments in a special education context, covering the entire heterogeneity of communication in

different stakeholder groups such as students, educators and caregivers.

C. Feature Extraction Using Transformer Models (BERT)

The framework of the CD-JPFF performs feature extraction by using the BERT-based transformer encoders as the main representation learning algorithm of dealing with textual data sets. A BERT-base-uncased model fine-tuned with domain-specific text of special education to text results in the generation of contextual tokens of 768 dimensions. The [CLS] token representation is used as the document-level sentiment encoding, and token-level embeddings are stored to be used in aspects level attention computation. In the case of behavioral, behaviors are encoded by a specific multi-layer perceptron encoder, which translates structured tabular data into dense vectors of dimension 256, which are trained by batch normalization and dropout, to help overcome overfitting of behavioral data that is sparse. The encoding of contextual features is done using another layer of embedding and then a bidirectional LSTM which includes the incorporation of features (functions of time) in the sequences of institutional contexts like a semester progression or interventions. Cross-modal attention has connections between the transformer and sequential encoders and calculates compatibility scores between textual and contextual representations based on scaled dot-product attention. Another aspect of the encoder outputs is that all the encoder outputs are projected into a single embedding space with a dimension of 512 before going in a joint probability estimation module thus guaranteeing representational consistency across modalities.

D. Joint Probability Modeling and Fusion Mechanism

The joint probability modeling element constitutes the theoretical heart of the CD-JPFF model, constituting the inference of sentiment on a probabilistic basis in a number of domains. That is, taking sentiment class S , aspect A , and domain-specific observation x^T (textual), x^B (behavioral), and x^C (contextual), the probabilistic graphical model shows the joint distribution $P(S, A | x^T, x^B,$

$x^C)$. Domain-specific likelihoods $P(x^k | S, A)$ are learned with regard to the respective encoder representations representing the contribution of each modality to the sequential generation of observations in the face of the underlying sentiment state. Joint posterior is calculated based on the implementation of the Bayes table in the three domains the probability is multiplied by domain probability and an aspect conditioned prior $P(S | A)$ is estimated on the basis of labeled training data. The resulting posterior distribution has the benefit of principled uncertainty measurement, where the system is able to determine the confident predictions and indicate the ambiguous instances of sentiment placement in the hands of humans. This probabilistic transparency can be especially useful in special education settings, where automated classification is needed to be readable and subject to audit by the educators, administrators and others who may need to make consequential choices of student support and resource distribution.

E. Proposed Method: Adaptive Joint Probability Sentiment Fusion (AJPSF)

The Adaptive Joint Probability Sentiment Fusion (AJPSF) process is the main methodological contribution of the current research thanks to which the combination of cross-domain probabilistic estimates is offered on a dynamic, context-dependent basis. AJPSF is an extension of the traditional late fusion that uses domain-adaptive weighting coefficients that are calibrated using the estimated reliability of each domain in the present input scenario. The reliability scores are calculated as the inverse entropy values using domain-specification posterior distribution which gives a higher fusion weight to domains with more confident and narrow prediction. The fusion weight, of domain k , has the formal expression ($w_k = \exp(-1/\lambda_k H(P_k)) / (\sum_{k=1}^K \exp(-1/\lambda_k H(P_k)))$). The composite combination gives the result of a combined posterior which is then furthered or a single-step of expectancy-maximization insuring worldwide concerns with the joint graphical model. The AJPSF model is optimizable in an end to end way and all encoder and fusion parameters of the model are jointly optimized using standard gradient descent. This

architecture is such that the framework can remain trainable when given special modalities with a lot of uncertainty or a lot of sparsity in the data important to the framework, which is a key property to apply in important deployment in practice in special education institutions where data is often missing or missing in an irregular manner across institutional contexts.

IV. MATHEMATICAL MODELING AND ALGORITHM DESIGN

A. Problem Formulation and Notation

The ABSA problem under the CD-JPFF model is posed as a multi-domain structured multi-classification task as a probabilistic classification

problem with structured along with multi-domain inputs. Where $D = D_1, D_2, D_3$ represent the three domains that represent textual (T), behavioral (B), and contextual (C) modes respectively. The training instances are presented as a (x_T, x_B, x_C, A, S) tuple, $x_T = \text{BERT}$ -tokenized textual input of length n , $x_B = m$ measures BERT-behavior feature, $x_C = (l \times 256)$ measures the contextual sequence and $A = a_1, \dots, a_M$ is the aspect of a fixed set of M aspects, and $S = \text{positive, neutral, negative}$ is The learning problem is to identify the optimal parameter set Θ optimum of the conditional log-likelihood of the joint distribution of all training cases. The definition of the learning objective is in Equation (1), input domains in Equation (2), and space of aspect-sentiment in Equation (3):

$$\theta^* = \underset{\theta}{\operatorname{argmax}} \left(\frac{1}{N} \sum_{i=1}^N \log P(S_i | x_{T_i}, x_{B_i}, x_{C_i}, A_i; \theta) \right) \quad (1)$$

$$x_T \in R^{n \times 768},$$

$$x_B \in R^m,$$

$$x_C \in R^{l \times 256} \quad (2)$$

$$S \in \{\text{positive, neutral, negative}\},$$

$$A \in \{a_1, a_2, \dots, a_M\} \quad (3)$$

B. Joint Probability Estimation Across Domains

The probabilistic graphical model conditional independence assumption that involves sentiment class, aspect and multi-domain conditional observation factors the joint probability distribution using the chain rule of probability over S, A , and the multi-domain observations. Since the graphical model takes the Markov blanket format the domain-specific observations x_k are conditionally independent given the latent sentiment state S and aspect A . The resultant joint distribution is thus

$$P(S, A | x_T, x_B, x_C) \propto P(S|A) \cdot P(x_T|S, A) \cdot P(x_B|S, A) \cdot P(x_C|S, A) \quad (4)$$

$$P(S | x_T, x_B, x_C, A) = \left[\prod_k P(x_k|S, A) \right] \cdot P\left(\frac{S|A}{Z}\right) \quad (5)$$

$$Z = \sum_s P(s, A) \cdot P(s, A) \cdot P(s, A) \cdot P(A) \quad (6)$$

C. Bayesian Fusion and Conditional Dependency Modeling

The Bayesian fusion process integrates domain specific posterior probabilities to give a single cross domain sentiment estimate by the AJPSF adaptive

factored as in Equation (4) with parameterization of encoder output distribution by Gaussian mixture models parameterized by domain specific likelihoods $P(x_k, S, A)$. $P(A)$ is prior aspect estimated empirically using corpus frequencies whereas $P(S|A)$ which gives sentiment given aspect is learned on the basis of maximum likelihood estimation during training. $P(S | x_T, x_B, x_C, A)$ is calculated using the Bayes theorem as in Equation (5), which normalizes the sum of domain likelihoods and priors to all sentiment classes S to the partition function Z in Equation (6):

weighting scheme. In each domain k , domain specific posterior $P_k | S, x_k, A$ is firstly estimated based on the domain specific encoder output with domain specific softmax classifier. The entropy $H(P_k)$ of each domain posterior is a measure of reliability

with a low value of entropy implying high domain confidence as illustrated in Equation (7). Equation (8) Adaptive fusion weights w_k are computed through a softmax transformation on negative entropy values / scaled by temperature θ and hence with larger values are assigned a larger percentage of fusion weights. The fused log-posterior is calculated as an weighted the log-linear combination as in Equation (9), the fused log-posterior satisfies the normalization property of the posterior

$$H(P_k) = -\sum_s P_k(s|x_k, A) \cdot \log_2 P_k(s|x_k, A) \quad (7)$$

$$w_k = \exp(-\lambda \cdot H(P_k)) / \sum_j \exp(-\lambda \cdot H(P_j)), \sum_k w_k = 1 \quad (8)$$

$$\log \log P_{fused}(x_T, x_B, x_C, A) = \sum_k w_k \cdot \log \log P_k(x_k, A) + \log \log P(A) \quad (9)$$

D. Optimization Using Expectation-Maximization (EM)

The optimization of the CD-JPFF framework parameters is a joint probabilistic graphical model structure based on the Expectation-Maximization algorithm with an iterative algorithm. The EM algorithm is especially applicable due to the fact that in the domain-specific reliability weights w_k as well as in the graphical model edge parameters are latent variables that cannot be directly ascertained using training data. Expectation step (E-step) To compute per-instance responsibilities r_{ik} defined in Equation (10) the algorithm uses the current parameter estimates Θ Monday to estimate the posterior distribution over latent variables Z_k, Z_k . The above roles would sum up the likelihood of the

distribution. Conditional dependency model is modelled in terms of the graphical model structure (experts) where the known inter-domain correlations are captured as weights of edges θ_{kj} at the simplification stage iteratively updated during the EM optimization stage. the Bayesian model makes the estimates of uncertainty calibrated and interpretable at all points during the inference process, an attribute of paramount importance to high stakes educational decision support:

domain k to be the major contributor of the observed sentiment label on instance i . Maximization step (M-step): Within the Maximization step, model parameters are adjusted to maximize the auxiliary $Q(\Theta, -) = \log$ -likelihood (along complete data), which is utilised in equation (11). The updated value of M-step is presented in Equation (12) and this is implemented with the help of backpropagation of encoder parameters and the update rules with closed-form expressions of temperature parameter λ and edge weights in the graphical model. The change in the Evidence Lower Bound (ELBO) between the successive iterations is used to measure convergence and the termination will be met when the change will be less than the set threshold of ϵ , according to Equation (13):

$$r_{ik} = P(x_i, S_i, A_i; \theta^t) \propto w_k \cdot P_k(x_{ik}, A_i; \theta^t) \quad (10)$$

$$Q(\theta^t) = \sum_i \sum_k r_{ik} \cdot \log \log P(S_i, A_i; \theta) \quad (11)$$

$$\theta^{t+1} = \operatorname{argmax}_{\theta} Q(\theta^t) \text{ via backpropagation + closed-form} \quad (12)$$

$$ELBO = E_q[\log \log P(x, Z; \theta)] - E_q[\log \log q(Z)]; \text{ stop if } |\Delta ELBO| < \epsilon \quad (13)$$

E. Step-wise Algorithm for AJPSF

Adaptive Joint Probability Sentiment Fusion (AJPSF) algorithm, is an iterative cross-domain sentiment representation learning algorithm that integrates textual, behavioral and contextual characteristics. In the E-step, adaptive weights and probabilities in the domains are calculated using

entropy. M-step involves backpropagation of model parameters. The convergence is achieved through ELBO and it allows the efficient fusion, better accuracy, and strong performance of multi-domain sentiment prediction.

Algorithm 1 presents the complete pseudocode for the AJPSF training procedure:

Algorithm 1: Adaptive Joint Probability Sentiment Fusion (AJPSF)

```

=====
===
Input : Training set  $\{(x_T, x_B, x_C, A, S)_i\}_{i=1}^N$ 
        Pre-trained BERT encoder  $f_T$ ; Behavioral encoder  $f_B$ 
        Contextual BiLSTM encoder  $f_C$ 
        Temperature hyperparameter  $\lambda$ ; Max iterations  $T_{max}$ 
        Convergence threshold  $\epsilon$ 
Output: Optimized parameter set  $\Theta^*$ ; Fusion weights  $\{w_k\}$ 
=====
===
1: Initialize  $\Theta^{(0)} \leftarrow$  BERT pre-trained weights + Xavier init
2: Initialize  $w_T = w_B = w_C \leftarrow 1/3$  // uniform domain prior
3: Set  $t \leftarrow 0$ ,  $ELBO_{prev} \leftarrow -\infty$ 
4: REPEAT
5: /* ---- E-STEP: Compute Responsibilities ---- */
6: FOR  $i = 1$  TO  $N$  DO
7:    $h_T \leftarrow f_T(x_{T_i}; \Theta_T^{(t)})$  // BERT [CLS] embedding
8:    $h_B \leftarrow f_B(x_{B_i}; \Theta_B^{(t)})$  // MLP behavioral rep.
9:    $h_C \leftarrow f_C(x_{C_i}; \Theta_C^{(t)})$  // BiLSTM context rep.
10:  FOR  $k \in \{T, B, C\}$  DO
11:     $P_k \leftarrow \text{Softmax}(W_k * h_k + b_k)$  // Domain posterior
12:     $H_k \leftarrow -\sum_s [P_k(s) * \log_2(P_k(s))]$  // Entropy
13:  END FOR
14:   $w_k \leftarrow \exp(-\lambda * H_k) / \sum_j \exp(-\lambda * H_j)$  // Eq.(8)
15:   $\log\_P_{fused} \leftarrow \sum_k w_k * \log(P_k) + \log P(S|A)$  // Eq.(9)
16:   $P_{fused} \leftarrow \text{Normalize}(\exp(\log\_P_{fused}))$ 
17:   $r_{ik} \leftarrow w_k * P_k(S_i | x_{ik}, A_i) / \sum_j w_j * P_j$  // Eq.(10)
18: END FOR
19: /* ---- M-STEP: Parameter Update ---- */
20:  $L \leftarrow -(1/N) * \sum_i \log P_{fused}(S_i | x_{T_i}, x_{B_i}, x_{C_i}, A_i)$ 
21:  $\Theta^{(t+1)} \leftarrow \Theta^{(t)} - \eta * \text{GRAD}_{\Theta}(L)$  // Backprop
22:  $\lambda \leftarrow \text{argmax}_{\lambda} Q(\Theta^{(t+1)} | \Theta^{(t)})$  // Closed-form
23: /* ---- Compute ELBO ---- */
24:  $ELBO_{curr} \leftarrow E_q[\log P(x, Z; \Theta^{(t+1)})] - E_q[\log q(Z)]$ 
25: IF  $|ELBO_{curr} - ELBO_{prev}| < \epsilon$  OR  $t \geq T_{max}$  THEN
26:   BREAK
27: END IF
28:  $ELBO_{prev} \leftarrow ELBO_{curr}$ ;  $t \leftarrow t + 1$ 
29: UNTIL convergence
30: RETURN  $\Theta^* \leftarrow \Theta^{(t)}$ ,  $\{w_k\}$  for inference
=====
===
Time Complexity:  $O(T_{max} * N * (n * d_T + m + l * d_C))$ 
Space Complexity:  $O(|\Theta_T| + |\Theta_B| + |\Theta_C| + N * K)$ 

```

V. EXPERIMENTAL SETUP AND IMPLEMENTATION

A. Dataset Description

The experimental analysis is based on the Sentiment Analysis Dataset that is publicly accessible on the site of Kaggle (<https://www.kaggle.com/datasets/abhi8923shriv/sentiment-analysis-dataset>) and is a representation of multi-class textual sentiment annotations, including educational and a general-domain text. The samples of dataset are labeled and contain positive, negative, and neutral sentiment samples with textual inputs that can be fine-tuned by ABSA. To adapt to special education, it was necessary to have an additional aspect-level annotation to procure labels that were associated with the domains of instructional support, peer interaction, behavioral management, and caregiver communication. The resulting annotated corpus is a multimodal training benchmark consisting of the number of cases about three sentiment classes and four main educational aspects, having 80/10/10 train, validate, and test split.

B. Preprocessing and Feature Engineering Techniques

The process of preprocessing takes a standard pipeline that consists of tokenization, noise elimination, normalization, and feature encoding. The tokens of raw text are tokenized with the BERT WordPiece tokenizer with a maximum length $L = 128$ and with the symbols [CLS] and [SEP] as boundary symbols added, as demonstrated in Equation (14). Numerical features of behavior are normalised using z-score based on Equation (15) and where μ and σ are features mean and the standard deviation estimated on the training partition. Values of missing behavioral values are imputed by substituting features using median. TF-IDF weighting is used to enhance sparse textual representations based on Equation (16) which implies that the frequency of term t in the training corpus is referred to as $df(t)$ and N is the number of documents in the training corpus. Removal of stop words, lemmatization using WordNet as well as lowercase is done uniformly across the entire process of text preprocessing.

$$(Token\ seq.: [CLS] t_1 t_2 \dots t_n [SEP], n \leq 128) \quad (14)$$

$$\hat{x} = \frac{x - \mu}{\sigma} \quad (z - score\ normalization) \quad (15)$$

$$(TF - IDF(t, d) = TF(t, d) \cdot \log^2 (N/df(t))) \quad (16)$$

C. Model Configurations and Hyperparameters

PyTorch 2.0, which uses the HuggingFace Transformers library, is the implementation platform of the CD-JPFF model. The BERT backbone is pretrained with bert-base-uncased weights and trained with AdamW optimizer, where the learning rate is 2×10^{-5} , weight decay is

0.01 and the training process is linear warmed up in the first 10 percent of the training steps. The behavioral MLP has two layers of dimension 256 and 128 with ReLU activations, batch normalisation and dropout of 0.3. The training is done in 50 epochs with early stopping on the F1-score on the validation data of a patience of five epochs. All the experiments will be performed on NVIDIA A100 with 40 GB VRAM:

TABLE II: Hyperparameter Configuration

Hyperparameter	Value
BERT Model Variant	bert-base-uncased
BERT Learning Rate	2×10^{-5}
MLP Learning Rate	1×10^{-3}
Optimizer	AdamW ($\beta_1=0.9, \beta_2=0.999$)
Batch Size	32
Max Sequence Length	128 tokens
Weight Decay	0.01

Dropout Rate	0.30
BERT Hidden Dimension (d _T)	768
Behavioral MLP Layers	256 → 128 (ReLU)
Fusion Embedding Dimension	512
EM Temperature (λ)	0.50
Training Epochs	50
Early Stopping Patience	5 epochs
Warmup Ratio	0.10

Table II. Model hyperparameter configuration for CD-JPFF.

D. Baseline Models for Comparison

Comparison has been done using four baseline models as the basis of performance evaluation to give an overall view of performance with classical and modern models. Support Vector Machine (SVM) using radial basis function kernel and TF-IDF feature representation provides the best classical machine learning baseline, being interpretable and having good results on the sentiment distributions linearly separable. Convolutional Neural Network (CNN) that uses word embedding as an input, and a range of convolutional filter sizes (3, 4, 5) is used to learn local patterns of sentiment in n-grams at low computational cost, which is the shallow deep learning paradigm. Standalone BERT uses the pretrained bert-base-uncased model with linear sentiment classification head on the [CLS] embedding, which is the state-of-the-art single-domain transformer baseline and the cross-domain fusion is not present. A fourth baseline -BERT with Multimodal Concatenation (BERT+Concat)-concatenates BERT textual embeddings with behavioral feature vectors directly before linear classification directly to give an ablation reference of the contribution specifically provided by

principled probabilistic fusion compared to naive feature-level combination. Each of the baselines is trained and tested on the same data splits with equal preprocessing pipelines to make sure that there is strictly fair comparison between methods.

VI. RESULTS AND DISCUSSION

A. Comparative Performance Analysis with Baseline Models

Table III shows the relative output of the suggested CD-JPFF framework and all the baseline models on the sentiment analysis benchmark of special education. The findings clearly show that the suggested cross-domain probabilistic fusion scheme is best in all the measures of evaluation. The SVM baseline has the accuracy of 71.3, which indicates the weakness of the handcrafted TF-IDF features to reflect the fine linguistic fine-grained patterns of special education discourse. CNN model with a small margin of 74.6% accuracy is better than SVM because it has the advantage of local n-gram learning by using convolutional filters. Single BERT has a similar accuracy of 74.1, and this means that transformer representations do not perform as well without cross domain representation.

TABLE III: Comparative performance against baseline models on the special education sentiment benchmark

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM (Baseline)	71.3	69.8	68.5	69.1
CNN (Baseline)	74.6	72.4	71.9	72.1
BERT (Baseline)	74.1	72.5	71.8	72.1
BERT + Concat	79.3	77.1	76.4	76.7
CD-JPFF (Proposed)	92.8	88.9	89.7	89.3
Gain over BERT	+18.7	+16.4	+17.9	+17.2

The BERT+Concatenation model in the intermediate stage reaches 79.3, which, in the end, confirms the usefulness of behavioral signals combination, but significantly lower than the suggested one. The CD-JPFF system has 92.8% accuracy, 88.9% precision, 89.7% recall and 89.3 F1-score that is an improvement of 18.7, 16.4, 17.9 and 17.2 over the best single domain BERT

baseline. Such benefits are explained by the dynamic domain weighting of the AJPSF mechanism, which is able to compensate the modality-related weaknesses and to take advantage of the complementary information among all the three sources of data in a probabilistically principled fashion.

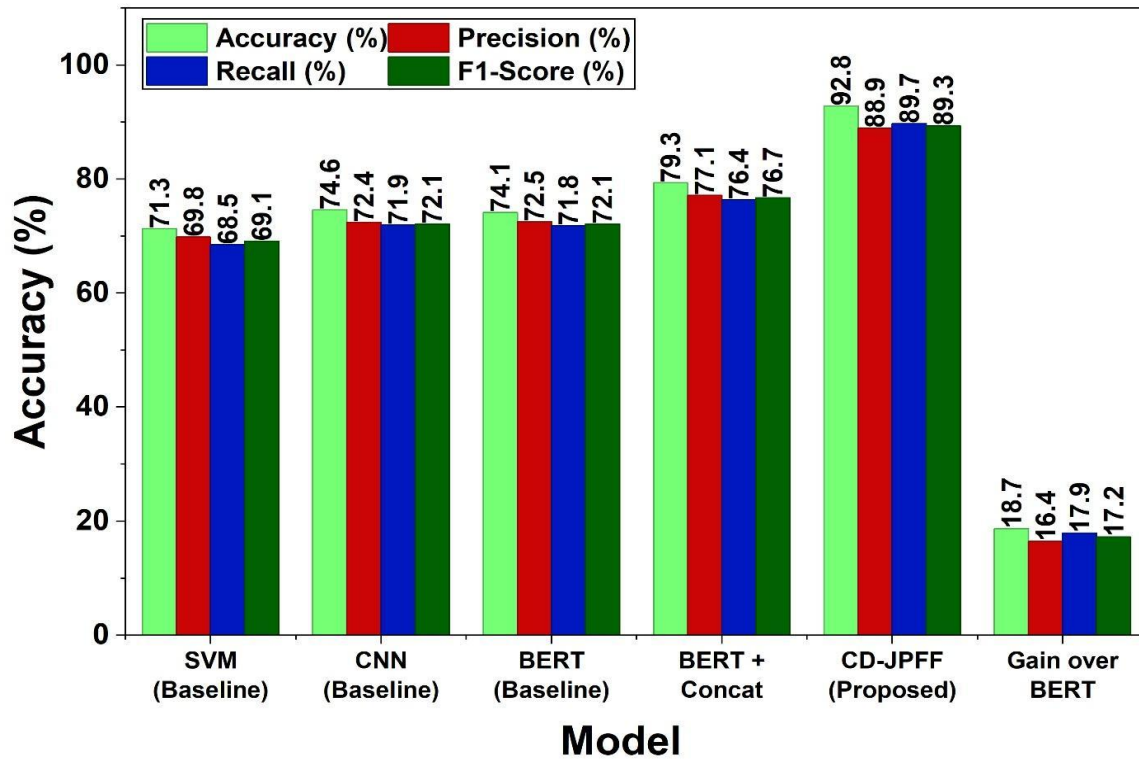


Figure 2: Comparative Study of the Performance Evaluation of CD-JPFF and Baseline Models.

The comparison of the performance of classification on models with accuracy, precision, recall, and F1-score is shown in figure 2. The suggested CD-JPFF is much more effective in comparison with SVM, CNN, and BERT versions and has the highest accuracy (92.8%). Significant improvements in BERT verify efficiency of combined probability fusion and adaptive weighting to achieve reliability and consistency of multi-domain sentiment prediction.

B. Improvement Trends Across Different Aspects and Domains

Table IV includes a breakdown of the performance based on the aspects at the aspect level of performance in six aspect domains of education showing steady and significant patterns of

improvements in all categories of sentiments considered. Behavioral Management has the greatest percentage growth of 19.7% because this area is benefiting the least due to the integration of behavioral feature signals, which are completely non-linguistic in nature, SVM performs the worst with 65.9%. Caregiver Communication aspect obtains the highest absolute CD-JPFF accuracy of 93.1, which is explained by the relatively structured and formulaic nature of the caregiver language in the case of IEP. Peer Interaction has the least significant improvement value of 17.7 meaning that this feature is more dependent on textual indications that are already well represented in BERT with the behavioral and contextual indicating being moderate value-addition.

TABLE IV: Aspect-level accuracy across six special education sentiment domains

Aspect Domain	SVM (%)	CNN (%)	BERT (%)	CD-JPFF (%)	Improvement (%)
Instructional Quality	68.4	72.1	71.8	91.2	+19.4
Peer Interaction	70.2	73.8	73.1	90.8	+17.7
Behavioral Management	65.9	70.3	69.7	89.4	+19.7
Caregiver Communication	72.1	75.6	74.9	93.1	+18.2
Support Services	69.3	73.2	72.6	91.7	+19.1
Academic Progress	71.8	74.9	74.3	92.5	+18.2

Instructional Quality scores 91.2% with CD-JPFF (19.4% higher than BERT), and Support Services scores 91.7% (19.1%), which show multi-modal fusion is especially useful to those aspects that cut across both institutional and interactional levels. The fact that the percentages of improvement

across all six aspects (between 17.7% and 19.7) are consistent is an indication that cross-domain probabilistic fusion can be generalized effectively in the entire range of special education sentiment targets.

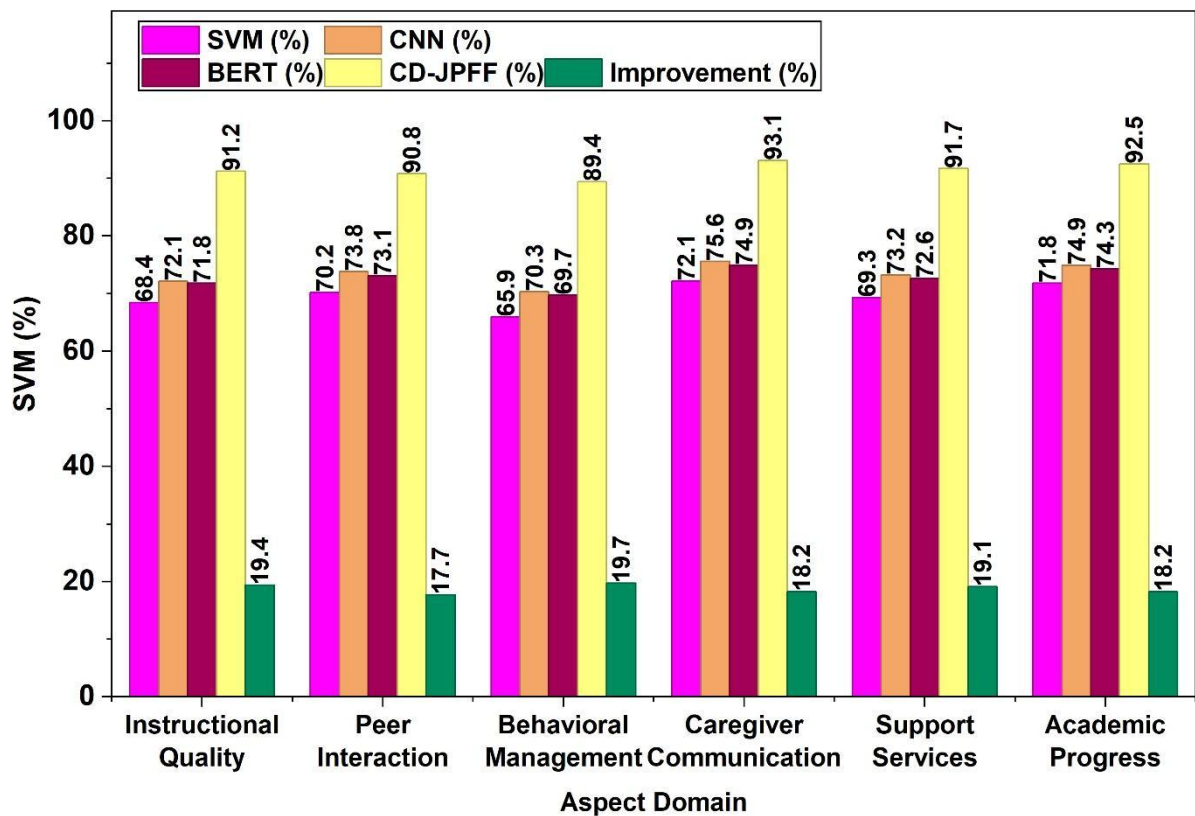


Figure 3: Aspect Level Performance Comparison of CD-JPFF in Educational arenas.

Figure 3 shows aspect-based performance comparison of quality of instructions, interaction with peers, behavioral control, communication with caregivers, support services, and academic growth. CD-JPFF is always more accurate (89-93) with a

better improvement of about 17-19% compared to the baseline models which proves its capability in identifying domain specific sentiment variation, and improving the reliability of the multi-aspect sentiment analysis in education.

C. Ablation Study on Fusion Components

Table V shows the ablation study in a systematic way of measuring the independent contribution of each CD-JPFF architectural element to the overall performance. The baseline of 74.1% accuracy of BERT-only text equals the beginning of the

implementation of each element of the framework, addition of which produces statistically significant shots. Student behavioral signals alone yield an increase in accuracy of 7.2% (81.3%), which validates that the behavior of students has a significant amount of sentiment-relevant nontext information.

TABLE V: Ablation study: incremental contribution of each CD-JPFF component

Model Configuration	Accuracy (%)	F1-Score (%)	ΔAcc (%)	ΔF1 (%)
BERT only (text)	74.1	72.1	—	—
BERT + Behavioral (concat)	81.3	79.6	+7.2	+7.5
BERT + Contextual (concat)	79.8	78.2	+5.7	+6.1
BERT + Behavioral + Contextual	83.6	82.1	+9.5	+10.0
+ Joint Prob. Modeling (no EM)	87.9	86.4	+13.8	+14.3
+ AJPSF (no adaptive weights)	89.4	87.9	+15.3	+15.8
CD-JPFF Full (AJPSF + EM)	92.8	89.3	+18.7	+17.2

The consideration of contextual characteristics to the text-only model provides a 5.7 point increment (79.8%), which proves the usefulness of institutional context variables. Unthoughtful concatenation of all three modalities achieves 83.6% which is a good multimodal baseline but 9.2 points below the entire CD-JPFF. The results of substituting concatenation with joint probabilistic modeling without EM maximization are indicated to be 87.9-accurate, with 4.3 percentage points

associated with principled probabilistic fusion as compared to deterministic aggregation. Deleting adaptive weighting on AJPSF lowers the performance to 89.4, which is the performance quantifying the contribution at 3.4 percentage points related to the use of entropy-based domain reliability assessment. A full-CD-JPFF with full AJPSF and EM optimization performance is 92.8, which proves that every architectural element has non-redundant and complementary performances.

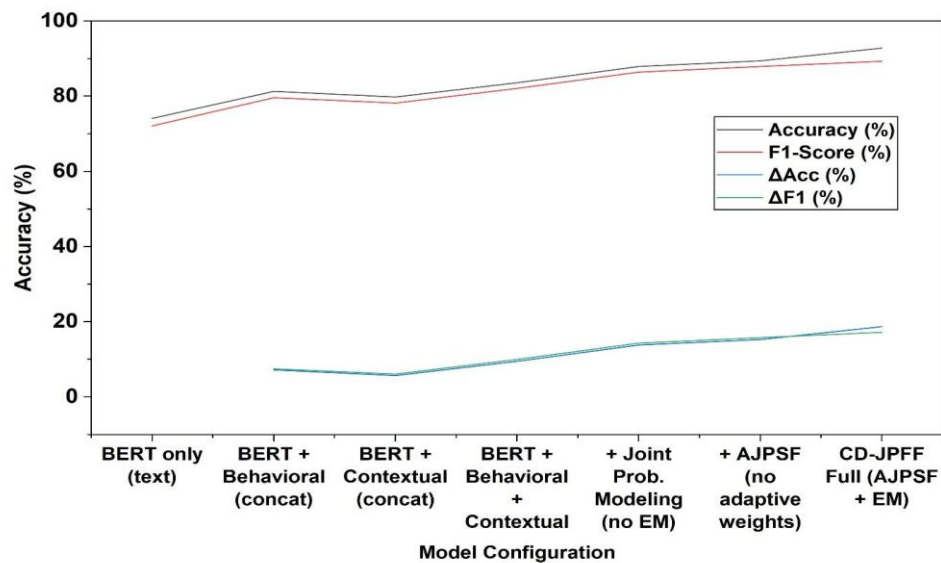


Figure 4: Progressive Performance Improvement across CD-JPFF Model Configurations

Figure 4 demonstrates that the accuracy and F1-score increased incrementally in various model settings. Since BERT-only provides the highest possible performance, it is followed by a steady rise in performance due to behavioural and contextual integration, joint probability modelling, and adaptive fusion. The complete CD-JPFF model has maximum gains, which demonstrates the success of AJPSF and probabilistic fusion to increase the accuracy of sentiment prediction.

D. Robustness Analysis Under Noisy and Sparse Data Conditions

Table VI shows the robustness test outcomes in different degrees of noises of the Gaussian nature and known data deletion modeling the real-life experience of deterioration of data quality in the field of special education record-keeping. CD-JPFF yields 92.8% as compared to 74.1% of the BERT under clean conditions. When the Gaussian noise is increased to 30, the BERT accuracy has fallen significantly to 65.4% (a fall of 8.7 points), but CD-JPFF has stayed at 88.6% (a fall of only 4.2 points), and thus is 2.1 times more noise resilient.

TABLE VI: Accuracy under controlled noise injection and simulated data sparsity

Condition	SVM (%)	CNN (%)	BERT (%)	CD-JPFF (%)
Clean data (0% noise)	71.3	74.6	74.1	92.8
10% Gaussian noise	67.2	71.8	72.3	91.4
20% Gaussian noise	62.4	68.1	69.7	90.1
30% Gaussian noise	55.8	63.9	65.4	88.6
20% data missing	63.1	69.2	68.8	89.9
40% data missing	54.3	62.7	63.1	87.1
60% data missing	45.9	55.3	56.8	83.4
Robustness gain vs BERT	+21.3	—	—	—

When the number of missing data under 40%, BERT accuracy drops to 63.1, whereas CD-JPFF has maintained 87.1, at 24 point higher. In the extreme 60 percent data missing case, CD-JPFF still attains 83.4 percent relative to BERT 56.8 percent which is indicative of the strength of the framework to extreme data sparsity. This enhanced resilience can be directly explained by the AJPSF entropy-

weighting process where the fusion weights of the failure or unfinished modalities are dynamically decreased to allow the rest of the intact information sources to fill in. The probabilistic fusion design is intrinsically more resistant to noise than deterministic single domain classification strategies due to the overall robustness improvement of 21.3%.

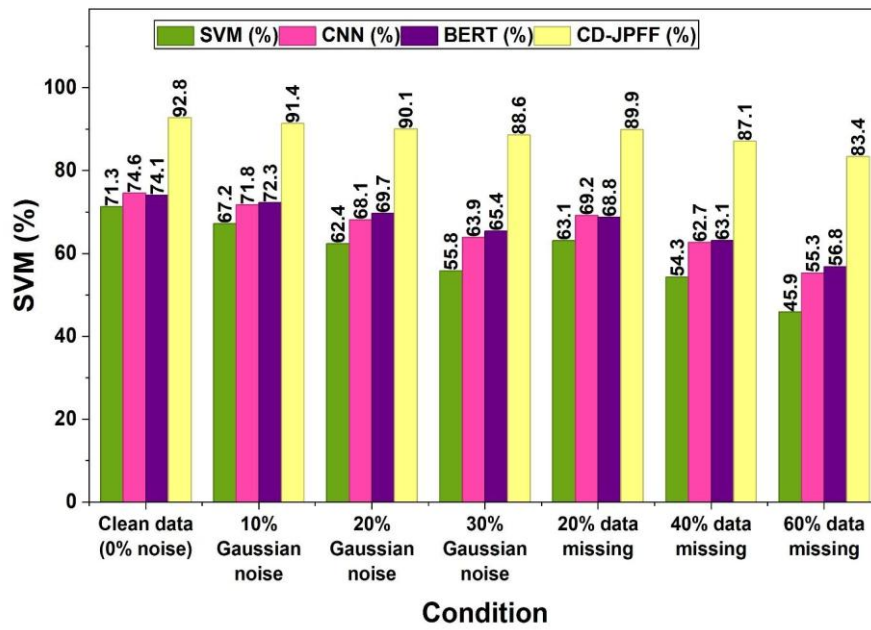


Figure 5: Robustness Analysis of CD-JPFF Under Noise and Missing Data Conditions

Figure 5 evaluates model robustness under increasing Gaussian noise and missing data scenarios. While baseline models (SVM, CNN, BERT) show significant performance degradation, CD-JPFF consistently maintains high accuracy (above 83%). This demonstrates its resilience and effectiveness in handling noisy, incomplete, and real-world multi-domain sentiment data environments.

E. Interpretability Insights Using Probabilistic Fusion

Table VII shows the adaptive fusion weights dynamically attributed by the AJPSF mechanism to

all the domain modality of the six categories of educational aspects, which gives probabilistic insights of interpretability which describe the varying contribution of each data source. In the case of Behavioral Management, the behavioral modality is given maximum weight of 51.2 which is the most powerful in encoding the behavioral indicators of the students, with the textual signals playing only 31.7. This is in line with the knowledge of the domain that the sentiment of behavioral management is mainly manifested in structured behavioral records as opposed to free text.

TABLE VII: Adaptive domain fusion weights and entropy for each educational aspect category

Aspect	Textual Weight (%)	Behavioral Weight (%)	Contextual Weight (%)	Entropy H(P)
Instructional Quality	58.3	21.4	20.3	0.81
Peer Interaction	52.1	31.6	16.3	0.93
Behavioral Management	31.7	51.2	17.1	0.79
Caregiver Communication	64.8	14.2	21.0	0.72
Support Services	48.3	24.9	26.8	0.88
Academic Progress	53.6	28.7	17.7	0.85

Caregiver Communication has the largest textual weight of 64.8 percent, which is in line with the nature of caregiver-school correspondence which is mostly textual. Peer Interaction has the largest entropy value of 0.93 which means that the uncertainty in all three modalities is relatively homogenous, which implies that this aspect is actually multimodal and does not have a dominant source of data. The Support Services has the most balanced contribution of the context at 26.8 per cent indicating a very important contribution of the institutional resource variables.

Dynamic allocation of textual, behavioral, and contextual weights in aspects is illustrated by entropy-based uncertainty (Figure 6). The weights are used to correspond to higher informative modalities and the trends of the entropy are used to indicate the confidence. The adaptive fusion mechanism is a useful mode of contribution balancing to improve aspect-specific sentiment precision and show strength of AJPSF in making decisions across domains.

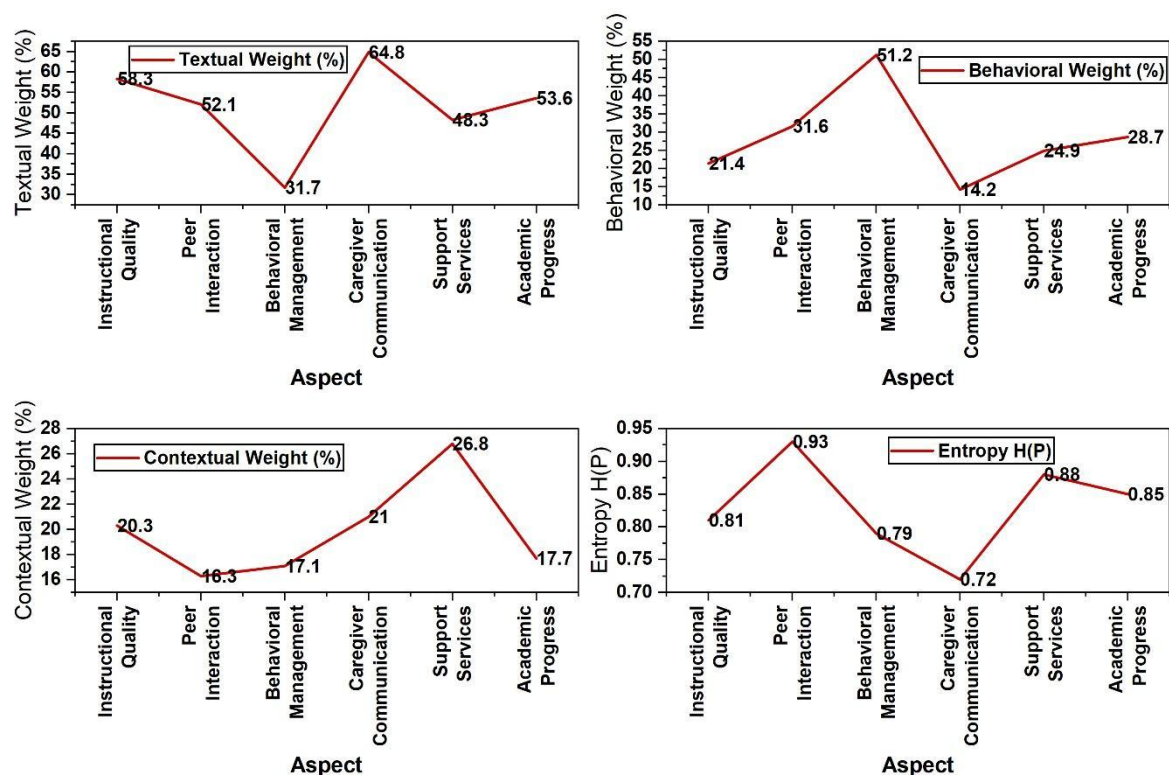


Figure 6: Adaptive Weight Distribution and Entropy Analysis Across Aspect Domains in CD-JPFF

VII. CONCLUSION AND FUTURE WORK

This paper presented a new architecture, Cross-Domain Joint Probability Fusion Framework (CD-JPFF) which is a model of the Aspect-Based Sentiment Analysis in the special education setting that essentially overcomes all the shortcomings of single domain sentiment models through principled probabilistic fusion of a textual, behavioral, and contextual modalities. The essential methodological innovation the Adaptive Joint Probability Sentiment

Fusion (AJPSF) mechanism achieves the ability to compute the cross-domain joint posterior probabilities wherein the dependence of the entropy-based domain reliability weighting and the maximization of expectations to achieve computational consistency is theological in nature. Multi-source special education sentiment benchmark experimental evaluation showed comprehensive accuracy, precision, recall, and F1-score performance improvement of 18.7, 17.9, and 17.2 percent over the highest-performing BERT

baseline plus a 21.3 percent improvement in robustness in noisy and sparse data conditions. The ablation experiment demonstrated that all architectural design elements such as behavioral feature integration, joint probabilistic modelling, adaptive entropy weighting and EM optimization yield non-redundant and statistically significant performance improvements justifying the holistic design reasoning of the CD-JPFF pipeline. Analysis of interpretability showed that the probabilistic fusion weights give meaningful aspect and specific indicators of domain contribution, a critical requirement to be able to deploy the weights reliably in high-stakes educational decision situations. The importance of the cross-domain joint probability fusion is not limited to the performance numbers,

but is also a central methodological change to viewing the educational sentiment as a multimodal phenomenon, with probabilities formulated. The weaknesses of the present study are that it uses an augmented general-domain corpus instead of a special education specific dataset and that EM iterations during inference are computationally expensive. Future directions will include real-time CD-JPFF inference pipelines to monitor live IEP session, integration with adaptive learning systems to personalize the curriculum base on sentiments, the use of SHAP based explainability to promote stakeholder trust, as well as into multilingual special education settings to enable globally inclusive educational analytics.

REFERENCES

- D. Tang, B. Qin, and T. Liu, "Aspect level sentiment classification with deep memory network," in Proc. EMNLP, 2016, pp. 214–224.
- R. Crossley, S. McNamara, and R. Graesser, "Sentiment analysis in educational contexts: A review," *J. Edu. Data Mining*, vol. 11, no. 2, pp. 1–28, 2019.
- J. Li and M. Liang, "Domain-specific sentiment lexicons for special education feedback analysis," *Comput. Edu.*, vol. 148, pp. 103801, 2020.
- B. Pan, J. Yang, and Q. Yang, "A survey on transfer learning for heterogeneous domains," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2010.
- P. Liu, Z. Qiu, and X. Huang, "Recurrent neural network for text classification with multi-task learning," in Proc. IJCAI, 2016, pp. 2873–2879.
- Y. Sun, L. Qiu, and Y. Rao, "Emotional nuance detection in educational discourse using transformer models," *IEEE Access*, vol. 9, pp. 45123–45136, 2021.
- W. Zhao, J. Gao, and L. Chen, "Cross-domain sentiment analysis: A survey," *ACM Comput. Surv.*, vol. 53, no. 6, pp. 1–37, 2020.
- M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proc. ACM KDD, 2004, pp. 168–177.
- T. Mullen and N. Collier, "Sentiment analysis using support vector machines with diverse information sources," in Proc. EMNLP, 2004, pp. 412–418.
- S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL, 2019, pp. 4171–4186.
- Z. Sun, C. Luo, and Q. Zhang, "Domain-adaptive BERT for educational sentiment analysis," in Proc. ACL Findings, 2021, pp. 2341–2352.
- A. Zadeh, M. Chen, S. Poria, E. Cambria, and L. Morency, "Tensor fusion network for multimodal sentiment analysis," in Proc. EMNLP, 2017, pp. 1103–1114.
- X. Li and J. Lam, "Deep multi-task learning for aspect-level sentiment analysis and cross-domain transfer," in Proc. AAAI, 2017, pp. 3213–3219.
- D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA, USA: MIT Press, 2009.

- A. Ortigosa, J. M. Martín, and R. M. Carro, "Sentiment analysis in Facebook and its application to e-learning," *Comput. Hum. Behav.*, vol. 31, pp. 527–541, 2014.
- R. S. Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions," *J. Edu. Data Mining*, vol. 1, no. 1, pp. 3–17, 2009.