Information Processing and Management 62 (2025) 103876

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/ipm

Leveraging sensory knowledge into Text-to-Text Transfer Transformer for enhanced emotion analysis

Qingqing Zhao [a], Yuhan Xia [b,*], Yunfei Long [b,*], Ge Xu [c], Jia Wang [d]

[a] Institute of Linguistics, Chinese Academy of Social Sciences, Beijing, 100732, China
[b] School of Computer Science and Electronic Engineering, University of Essex, Colchester, CO4 3SQ, UK
[c] College of Computer and Control Engineering, Minjiang University, Fuzhou, 350108, China
[d] Department of Intelligent Science, Xi'an Jiaotong-Liverpool University, Suzhou, 215123, China

ARTICLE INFO

Keywords:
Emotion analysis
Sensory knowledge
Attention mechanism
Pre-trained language model

ABSTRACT

This study proposes an innovative model (i.e., SensoryT5), which integrates sensory knowledge into the T5 (Text-to-Text Transfer Transformer) framework for emotion classification tasks. By embedding sensory knowledge within the T5 model's attention mechanism, SensoryT5 not only enhances the model's contextual understanding but also elevates its sensitivity to the nuanced interplay between sensory information and emotional states. Experiments on four emotion classification datasets, three sarcasm classification datasets one subjectivity analysis dataset, and one opinion classification dataset (ranging from binary to 32-class tasks) demonstrate that our model outperforms state-of-the-art baseline models (including the baseline T5 model) significantly. Specifically, SensoryT5 achieves a maximal improvement of 3.0% in both the accuracy and the F1 score for emotion classification. In sarcasm classification tasks, the model surpasses the baseline models by the maximal increase of 1.2% in accuracy and 1.1% in the F1 score. Furthermore, SensoryT5 continues to demonstrate its superior performances for both subjectivity analysis and opinion classification, with increases in ACC and the F1 score by 0.6% for the subjectivity analysis task and increases in ACC by 0.4% and the F1 score by 0.6% for the opinion classification task, when compared to the second-best models. These improvements underscore the significant potential of leveraging cognitive resources to deepen NLP models' comprehension of emotional nuances and suggest an interdisciplinary research between the areas of NLP and neuro-cognitive science.

1. Introduction

1.1. Emotion analysis

Emotion analysis (EA), enabling machines to infer and understand human emotions, has been demonstrated with diverse applications, such as affective computing, digital healthcare, education, and so on Khare, Blanes-Vidal, Nadimi, and Acharya (2024) and Mitra, Nie, and Azemi (2024). In recent decades, various sensory resources from diverse sources and modalities, such as physical signals (e.g., speech and facial expression) and physiological signals (e.g., electroencephalogram, electrocardiogram, galvanic skin response, and eye tracking), have been extensively used for automatic emotion recognition, which have improved the models for

* Corresponding authors.
E-mail addresses: zhaoqq@cass.org.cn (Q. Zhao), yx23989@essex.ac.uk (Y. Xia), yl20051@essex.ac.uk (Y. Long), xuge@pku.edu.cn (G. Xu), Jia.Wang02@xjtlu.edu.cn (J. Wang).

https://doi.org/10.1016/j.ipm.2024.103876
Received 24 April 2024; Received in revised form 26 August 2024; Accepted 27 August 2024
Available online 4 September 2024

emotion recognition greatly (Fan et al., 2023; Raheel, Majid, Alnowami, & Anwar, 2020; Rodriguez et al., 2022; Skaramagkas et al., 2023; Tuncer, Dogan, & Acharya, 2021; Zhong, Wang, & Miao, 2022).

However, in Natural Language Processing (NLP), the task of EA has received less notable results in recent years. One of the reasons is that different from the sentiment analysis (SA) which provides a coarse-grained category of polarities including the positive, negative, or neutral values (Long, Xiang, Lu, Huang, & Li, 2019), EA needs to paint a more detailed picture. That is, the task of EA not only distinguishes between basic polarities, but also identifies nuanced emotions such as joy, anger, sadness, surprise, and among others (Ekman, 1992). In addition, when the medium only involves the textual content, it is not easy to distinguish closely-related emotions like "annoyance" and "anger" or "sadness" and "grief", as illustrated in examples (1) and (2) respectively based on the GoEmotions dataset (Demszky et al., 2020). These facts suggest that a discerning approach is required for emotion analysis.

(1) *"I feel like one day you're in for the shock of your life"*.

(2) *"If that mask could talk it would beg someone to end its suffering"*.

It should also be noted that unlike emotion recognition in non-textual media using extensive sensory resources (Khare et al., 2024), EA in NLP has paid little attention to sensory knowledge, although the intimate relationship between emotion and sensory perceptions has been verified repeatedly in various disciplines. Specifically, from a neurological perspective, emotion and sensory information are processed in an overlapping neural region, i.e., the amygdala (Šimić et al., 2021). Based on a meta-analysis of affective experiences across various sensory modalities, research by Satpute et al. (2015) has also found that affective experiences are constructed from activities distributed across both limbic and sensory cortical regions, which highlights the integral role of sensory processing in shaping affective experiences. Shifting the lens to psychology, emotion and sensation are intertwined (Zadra & Clore, 2011). For example, the sense of taste shows an inherent link with reward and aversion mechanisms, such as sucrose being perceived as sweet and desirable, whereas quinine being recognized as bitter and repulsive (Yamamoto, 2008). In addition, emotion as a kind of interoception forms an indispensable part of human sensations (Connell, Lynott, & Banks, 2018; Lynott, Connell, Brysbaert, Brand, & Carney, 2020). The neural and psychological connections between emotion and sensory experiences form the foundation for that infusing sensory resources into machine learning models can consistently contribute to automatic emotion recognition in non-textual media (Khare et al., 2024).

It is also important to note that the intimate relations between emotion and sensory experiences are also encoded in the natural language, due to the embodied cognition of human beings (Gibbs, 2005; Meteyard, Cuadrado, Bahrami, & Vigliocco, 2012; Wilson, 2002; Wilson & Golonka, 2013). That is, the patterns of how humans experience the world and interact with surrounding environments neurologically and cognitively are simulated in the language. Take describing emotions in the language for example. The conceptual metaphor **EMOTION IS PERCEPTION** is grounded in abundant language usages, which shows that the vocabulary related to human sensory experiences are fruitful sources for verbalizing emotions (e.g., *"sweet"* generally for happiness and *"bitter"* generally for sadness) (Kövecses, 2019; Lakoff & Johnson, 1980; Müller, Nagels, & Kauschke, 2021). Furthermore, it has been attested that people more frequently use sensory terms (e.g., *"warm"*, *"sour"*, *"bright"*, etc.), instead of literal emotion terms (e.g., *"happy"*, *"sad"*, *"angry"*, etc.), to convey emotions figuratively in the language (Fainsilber & Ortony, 1987; Lee, 2018).

Thus, emotion shows an intimate relationship with sensory perceptions both in perceptually and linguistically. However, traditional approaches in NLP often treat sensory perception and emotion classification as distinct fields, overlooking the potential usefulness of sensory knowledge in the emotion classification task.

1.2. Aims of this study

This study aims to propose a SensoryT5 model, designed to infuse sensory knowledge encoded in the language into neural architectures for a comprehensive approach to the automatic emotion analysis in NLP. Specifically, sensory knowledge is infused into T5 (The Text-to-Text Transfer Transformer Model) (Raffel et al., 2020) using an adapter approach built upon attention mechanisms. In addition, the contextual and sensory information learning branches are amalgamated within a unified loss function to facilitate joint training. The main contributions of our work can be summarized as follows:

- We propose an innovative architecture (i.e., SensoryT5), which enhances the transformer-based fine-grained emotion classification model by seamlessly embedding sensory knowledge. Marking one of the pioneering endeavors, SensoryT5 is adapted to harmonizing both the nuances of contextual attention and the intricacies of sensory information-based attention.
- Experiments on multiple publicly available datasets for fine-grained emotion classifications show that our approach improves the efficacy of the pre-existing model considerably and outperforms the state-of-the-art baseline models. In addition, our proposed model also achieves state-of-the-art results on other intricately complex tasks for affective analysis, including sarcasm classification, subjectivity analysis, and opinion extraction.
- Apart from the superiority and consistency of our SensoryT5 model over baseline models, infusing sensory knowledge for word representation in the model can enhance the interpretability of the results. This would not only benefit the explainability of the model and its applications but also further our understanding of the emotional and cognitive processes of humans.
- SensoryT5 leverages sensory knowledge within transformer text classification frameworks, contributing to the ongoing efforts to incorporate neuro-cognitive data in NLP tasks. That is, our work illuminates the potential of sensory information in refining emotion analysis, carving fresh prospects for exploration within affective computing in NLP. In addition, this study underscores the value of cognition-anchored resources in sculpting attention models, which also encourages continued interdisciplinary dialogue and research between the domains of NLP and neuro-cognitive science.

2. Related work

2.1. Emotion analysis

Emotion analysis has evolved significantly over time, transitioning from rule-based methods to advanced pre-trained language models (PLMs) and large language models (LLMs). Initially, the research on EA mainly centres on learning features through the use of neural networks such as Convolutional Neural Networks (CNN) (Kim, 2014), Recurrent Neural Networks (RNN), and Long Short-Term Memory networks (LSTM) (Hasim Sak & Beaufays, 2014), focusing primarily on syntactic parsing (Chen, 2022). Over recent years, PLMs and LLMs have achieved marked advancements. Noteworthy developments include models like BERT (Devlin, Chang, Lee, & Toutanova, 2018), RoBERTa (Liu et al., 2019), GPT-3 (Brown et al., 2020), T5 (Raffel et al., 2020), PaLM (Chowdhery et al., 2022), LLaMA (Touvron et al., 2023), and ChatGPT (OpenAI, 2023). These models, through rigorous pre-training on vast text corpora using self-supervised learning, have the ability to autonomously generate intricate representations. The capability has significantly advanced the field, setting new benchmarks in different tasks, notably in sentiment and emotion analysis (Devlin et al., 2018; Lu et al., 2023; Peng et al., 2024; Přibáň, Šmíd, Steinberger, & Mištera, 2024; Zhang, Deng, Liu, Pan, & Bing, 2023). Among the models, T5 (Raffel et al., 2020) stands out along the similar sized models due to its innovative text-to-text transfer approach, in which every NLP challenge is remodelled as a text-to-text problem.

Despite the considerable improvements made with PLMs and LLMs, some research gaps remain. For example, although the models possess sophisticated neural architectures capable of discerning patterns from immense text datasets, they have often overlooked the contribution of the integration of cognitive resources to the improvements of performances of the models. Nevertheless, recent explorations in the field suggest that integration of sensory resources with LLMs can potentially elevate their performances, nudging them closer to approaching human-like comprehension (Khare et al., 2024). For example, the study conducted by Long et al. (2019) has demonstrated the efficacy of utilizing eye-tracking data in improving sentiment analysis outcomes significantly. A study by Yan, Zhang, and Zhang (2024) has shown that combining electroencephalogram and eye-tracking signals can markedly improve Automatic Keyphrase Extraction (AKE) from microblogs. Similarly, a study by Li, Zhang, Wang and Gao (2021) demonstrated that a cognitive brain model based on attention neural networks significantly improves multimodal sentiment analysis. Additionally, Chen, Huang and Xue (2021) showed that a bilateral-brain-like cognitive network enhances aspect-level sentiment analysis by mimicking human cognitive processes. In addition, the evolution of models like ChatGPT to include multimodal capabilities, such as voice and image processing, marks a significant shift in the field. The advancement enriches the tapestry of human–machine interaction by allowing a more intuitive and context-rich interaction, challenging the traditional cognitive focus of these models (Nosta, 2023). Thus, the synergetic integration of LLMs with cognitive architectures is a promising direction. That is, by creating and dynamically updating cognitive models, the integration would enhance the understanding and interpretation of complex knowledge about entities and their relationships, thereby fortifying the emotion analysis capabilities of LLMs (Romero, Zimmerman, Steinfeld, & Tomasic, 2023).

This study presumes that integrating sensory resources and cognitive architectures with LLMs such as T5 would also significantly augment their efficacies in emotion analysis, marking a pivotal advancement in the field. Thus, our proposed SensoryT5 model is designed to synergize the strengths of T5 and augment it with sensory knowledge, enabling a deeper and more nuanced understanding of human emotions.

2.2. Sensory resources: Lancaster norms

In recent years, there has been an emergent trend that neuro-cognitive data and computational approaches are synergized in NLP studies. The interdisciplinary synergy unlocks new dimensions in understanding language, perception, and cognition for human beings. Most of the studies focus on using neuro-cognitive data for metaphor detection. For instance, Chen, Hai et al. (2021) incorporated the brain measurement data for modelling word embedding to identify metaphorical usages. Wan, Su, Ahrens, and Huang (2023) demonstrated the superiority of neural networks for metaphor detection by leveraging sensorimotor knowledge. These studies collectively underscore a broader shift in the field towards a more integrated approach to NLP. By weaving in neuro-cognitive data, researchers are equipping computational models with a richer and more intricate understanding of language and cognition, which are often overlooked by traditional data-driven methods.

Given the intimate connection between emotion and sensory experiences as demonstrated in various studies reviewed in the last section, this study assumes that a cognitively and linguistically motivated representation of words in text based on sensory knowledge would improve the performance of computational models for emotion analysis. That is not only because sensory inputs are crucial sources for emotions, but also because emotions are part of sensory perceptions for human beings (Connell et al., 2018; Lynott et al., 2020). In addition, emotions are overwhelmingly expressed in terms of sensory words (Fainsilber & Ortony, 1987; Lee, 2018). This would also facilitate the computational models that capture the correlation between emotion and sensory experiences by infusing sensory knowledge encoded in linguistic items for emotion analysis.

This study utilizes (Lynott et al., 2020)'s sensorimotor norms which encompass the metrics of sensorimotor strengths (ranging from 0 to 5) of 39,707 English words spanning six perceptual domains including touch, taste, smell, vision, hearing, and interoception, as well as five action effectors including mouth/throat, hand/arm, foot/leg, head (excluding mouth/throat), and torso. Lynott et al. (2020)'s sensorimotor norms (named "Lancaster norms" hereinafter) were compiled by following the sensory rating task proposed by Lynott and Connell (2009, 2013), which asked participants to rate the extent to which the meaning of a lexical item is based on sensory perceptions through the six sensory modalities and the five action effectors. Thus, the

Table 1
The sensory ratings of six sample words in the Lancaster norms.

Words	Touch	Taste	Smell	Vision	Hearing	Interoception
Soft	4.526	0.368	0.316	1.947	0.684	0.842
Flavour	0.219	4.938	1.719	0.344	0.094	1.094
Incense	1.412	0.294	5.000	2.824	0.176	0.295
Blue	0.150	0.000	0.000	4.450	0.250	0.500
Noisy	0.244	0.080	0.090	1.006	4.752	0.920
Headache	0.650	0.000	0.000	0.400	0.400	4.900

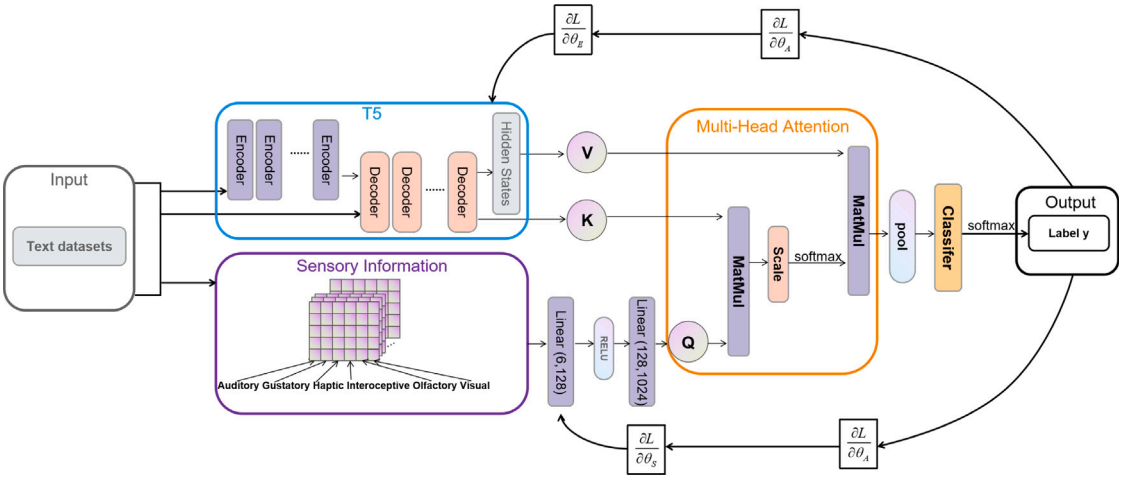

Fig. 1. An overview of SensoryT5. The blue box shows a T5 process of deep learning, and the purple box describes sensory information quantified and passed into T5. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

norms are language-specific lexical properties representing the correlation between conceptualized lexical meanings and sensory modalities/action effectors. As there is little work reported to show the correlation between emotion and sensory action effectors, this study only exploits the perceptual data in the six sensory modalities. Table 1 shows the perceptual ratings of six sample words in the Lancaster norms.

This study introduces SensoryT5, a model designed to construct sensory vectors using the Lancaster norms. These vectors are embedded into the T5's decoder mechanism through an auxiliary attention layer. Positioned after the decoder, this sensory-centric attention layer synergizes with the decoder's output, creating an enriched representation filled with sensory knowledge of words in the text. Consequently, SensoryT5 is adapted to simultaneously discerning contextual cues and sensory knowledge, facilitating a potent alignment of sensory nuances with contextual intelligence. This integration would enhance the model's efficacy in emotion analysis.

3. Our SensoryT5 model

The overarching structure of the proposed SensoryT5 model is depicted in Fig. 1. Specifically, sensory knowledge is infused into T5 (Raffel et al., 2020) using an adapter approach built upon attention mechanisms. Besides, the contextual and sensory information learning branches are amalgamated within a unified loss function for joint training.

Our selection of T5 as the foundational model is because tasks involving sensory elements are unlikely to be confined to emotion classification exclusively. Tasks such as question-answering and text generation may also benefit from the integration of sensory information, as incorporating sensory aspects into AI tasks, especially in text generation, echoes the longstanding aim of achieving more human-like artificial intelligence (Duéñez-Guzmán, Sadedin, Wang, McKee, & Leibo, 2023; Khanam, Tanweer, Khalid, & Rosaci, 2019). T5 distinguishes itself with its innovative text-to-text transfer approach, where each NLP challenge is remodelled as a text-to-text task. This crucial characteristic of the T5 model would offer substantial versatility for a variety of applications. Thus, leveraging T5's adaptable structure, SensoryT5 can be applied to a broad spectrum of tasks, particularly those where sensory information is crucial. For example, in question-answering or text generation, SensoryT5 has the potential to handle complex scenarios more effectively, producing outcomes that are enriched with sensory details, thereby elevating the overall quality.

3.1. Preliminaries

Despite the large size of Lancaster norms, there are still out-of-vocabulary words. Following the method proposed by Li, Lu, Long, and Gui (2017), we use a word embedding model to regressively predict the sensory values of unknown words, aiming to

obtain the sensory values for the out-of-vocabulary words. The details of predicting the sensory values for the out-of-vocabulary words will be given in Section 4.2.

Inputs and outputs. The objective of emotion analysis is to determine and categorize emotions for a piece of text following a defined label schema. Let D denote a collection of documents for emotion classification. Each document $d \in D$ is first tokenized into a word sequence with the maximum length n, then the word embeddings w_i of these sequences are jointly employed to represent the document $d = w_1, w_2, \ldots, w_i, \ldots, w_n (i \in 1, 2, \ldots, n)$.

3.2. The core attention mechanism in T5

The word embeddings of these sequences $d = w_1, w_2, \ldots, w_i, \ldots, w_n (i \in 1, 2, \ldots, n)$ first enter the T5 model. Each layer of the encoder and decoder has a series of multi-head attention units. The multi-head attention mechanism for the final decoder layer can be represented using the following equation:

$$\begin{aligned} V_d &= \text{MultiHead}(Q_0, K_0, V_0) \\ &= [\text{head}_1, \text{head}_2, \ldots, \text{head}_i] W_0 \end{aligned} \tag{1}$$

where each head is computed as:

$$\begin{aligned} \text{head}_i &= \text{Attention}(Q_0 W_i^Q, K_0 W_i^K, V_0 W_i^V) \\ &= \text{softmax}\left(\frac{(Q_0 W_i^Q)(K_0 W_i^K)^T}{\sqrt{d_k}}\right) V_0 W_i^V \end{aligned} \tag{2}$$

W_i^Q, W_i^K, and W_i^V are weight matrices that are learned during the training process. They are used to project the input queries (Q), keys (K), and values (V) to different sub-spaces. Q_0, K_0, and V_0 are derived from the output of the penultimate decoder layer. Additionally, following the common practice for text classification with the T5 model, we employ a zero-padding vector as the sole input for the decoder. The result V_d is the output of the T5 decoder, imbued with context-aware attention. Both V_d and K_0 will be utilized for the integration with sensory knowledge.

3.3. Sensory information transformation for T5 integration

We project the perceptual ratings of words in Lancaster norms and the predicted sensory values of the out-of-vocabulary words into a word vector space. Each word is linked with a six-dimensional vector representing sensory scores across six perceptual modalities (haptic, gustatory, olfactory, visual, auditory, and interoceptive dimensions). For a word w, its sensory vector is denoted as $s(w) = [s_1, s_2, \ldots, s_6]$.

To enable effective integration into the T5 model, we use two linear transformations to map the sensory vectors to the same dimension as the T5's word embeddings. Given a T5 model with an embedding dimension of 1024, the transformation process can be formally described as:

$$\mathbf{h}_1 = \text{ReLU}(\mathbf{W}_1 s(w) + \mathbf{b}_1) \tag{3}$$

$$s'(w) = \mathbf{W}_2 \mathbf{h}_1 + \mathbf{b}_2 \tag{4}$$

where $\mathbf{W}_1 : \mathbf{R}^6 \to \mathbf{R}^{128}$ and $\mathbf{W}_2 : \mathbf{R}^{128} \to \mathbf{R}^{1024}$ are two linear transformation matrices and $\mathbf{b}_1$ and $\mathbf{b}_2$ are the respective bias terms. The shapes of the two weight matrices W_1 and W_2 are $(6, 128)$ and $(128, 1024)$ respectively. The output h_1 of the first linear layer is a vector of shape $(1, 128)$, and the output $s'(w)$ of the second linear layer is a vector of shape $(1, 1024)$. After the transformation, the sensory vector $s'(w)$ is projected into the same semantic space as the features generated by the T5 model. The output vector $s'(w)$ with V_d and K_d from the T5 model will be applied in Section 3.4 for infusing sensory knowledge into the T5 model.

3.4. Sensory attention mechanism in SensoryT5

The sensory vector $s'(w)$ generated by the sensory vector transformation is used as the query in the attention mechanism of the sensory adapter, substituting the query vector Q in the T5 model. The sensory adapter performs the attention calculation as follows:

$$\begin{aligned} A_d &= \text{MultiHead}(s'(w), K_0, V_d) \\ &= [a_1, a_2, \ldots, a_i] W_d \end{aligned} \tag{5}$$

where each head is computed as:

$$\begin{aligned} a_i &= \text{Attention}(s'(w) W_i^Q, K_0 W_i^K, V_d W_i^V) \\ &= \text{Softmax}\left(\frac{(s'(w) W_i^Q)(K_0 W_i^K)^T}{\sqrt{d_k}}\right) V_d W_i^V \end{aligned} \tag{6}$$

Once the output $A_d = a_1, a_2, \ldots, a_n$ of the sensory adapter is obtained, we apply dropout and pooling operations to form a final representation P_d, which is then used as the input to the classification layer.

$$P_d = \text{Dropout}(\text{Pool}(A_d)) \tag{7}$$

The pooled representation P_d is then fed into the classifier of the T5 model.

$$C_d = \text{Softmax}(Linear(\text{Dropout}(P_d))) \tag{8}$$

C_d is a probability distribution vector. The class with the highest probability is selected as the predicted label, denoted as y.

The first step of the back-propagation process involves computing the gradient of the loss function with respect to the parameters of the sensory attention adapter. Θ_{A} represents the parameters of the sensory attention layer, and A_d represents the output of the sensory T5. The computed gradient is used to update the parameters of the attention layer, enhancing its capacity to integrate sensory information into the T5 model. This is represented as follows:

$$\frac{\partial \mathcal{L}}{\partial \Theta_{\text{A}}} = \frac{\partial \mathcal{L}}{\partial A_d} \cdot \frac{\partial A_d}{\partial \Theta_{\text{A}}} \tag{9}$$

After the gradients for the sensory attention mechanism have been computed, we then compute the gradients for the parameters of the final layer of T5, denoted as Θ_{E}.

$$\frac{\partial \mathcal{L}}{\partial \Theta_{\text{E}}} = \frac{\partial \mathcal{L}}{\partial V_d} \cdot \frac{\partial V_d}{\partial \Theta_{\text{E}}} \tag{10}$$

Finally, the gradients for the sensory information transformation, denoted as Θ_{S}, are computed as follows:

$$\frac{\partial \mathcal{L}}{\partial \Theta_{\text{S}}} = \frac{\partial \mathcal{L}}{\partial s'(w)} \cdot \frac{\partial s'(w)}{\partial \Theta_{\text{S}}} \tag{11}$$

Here, Θ_{S} represents the parameters of the sensory information transformation component, which includes the weights and biases of the two linear layers, and $s'(w)$ represents the output of this component. The calculated gradient is used to update the parameters of the sensory information transformation to improve its ability to capture and model the sensory information.

The loss function applied in this model is the cross-entropy loss, which is commonly used for classification tasks. Cross-entropy loss measures the performance of a classification model whose output is a probability value between 0 and 1. The loss increases as the predicted probability diverges from the actual label. Mathematically, for a single sample, the cross-entropy loss $\mathcal{L}$ is defined as:

$$\mathcal{L} = -\sum_{i=1}^{C} y_i \log(p_i) \tag{12}$$

where C is the number of classes, y_i is the binary indicator (0 or 1) if class label i is the correct classification for the given observation, and p_i is the predicted probability for class i. For multi-class classification, this loss function effectively penalizes the model for deviating from the true distribution of the classes. By minimizing the cross-entropy loss during training, the model learns to predict probabilities that are close to the true class labels. This loss function is particularly suitable for the current model as it deals with classification tasks and helps in adjusting the model parameters to achieve better classification accuracy.

Through these calculations, we are able to update the parameters of the sensory attention mechanism, the T5 model, and the sensory information transformation component.

4. Experiments

4.1. Datasets

To test the effectiveness of the SensoryT5 model, we selected datasets from two domains: emotion classification and sarcasm classification.

Below are the details of the four selected emotion classification datasets.

- **GoEmotions** (Demszky et al., 2020) is a benchmark dataset drawn from Reddit. Each comment was annotated with one of 27 distinct emotion categories or marked as neutral. In alignment with previous research (Suresh & Ong, 2021), our study only utilizes the single-labelled samples, with those categorized as neutral excluded, to ensure consistency and comparability of the results in emotion analysis.
- **Empathetic Dialogues (EmD)** (Rashkin, Smith, Li, & Boureau, 2019) presents a rich collection of dialogues, in which each conversation was uniquely labelled with one of 32 emotions. This dataset is utilized by feeding the situational context of each dialogue into models, effectively anchoring the analysis on the predefined emotion label, thus facilitating a nuanced understanding of empathetic responses in conversational settings.
- **ISEAR** (International Survey on Emotion Antecedents and Reactions) (Scherer & Wallbott, 1994) comprises a diverse array of sentences, each encapsulating a unique emotional experience. These sentences were categorized into one of seven primary emotion categories, providing a foundational resource for exploring the spectrum of human emotional responses in various contexts.

Table 2
Statistics of the selected datasets. "N_{train}" and "N_{test}" represent the number of instances in the training and testing sets respectively. "L" stands for the average text length within the dataset, and "C" indicates the number of categories.

	Dataset	N_{train}	N_{test}	L	C
Emotion	GoEmotions	23,485	2984	12	27
	EmD	19,533	2547	18	32
	ISEAR	4599	1534	22	7
	EmoInt	3612	3141	16	4
Sarcasm	SemEval 2018	3834	784	13	2
	Ghosh	46,070	3742	16	2
	IAC-V2	4179	465	48	2
Subjectivity	SUBJ	9000	1000	21	2
Opinion	PC	32,097	13,759	8	2

- **EmoInt** (Mohammad & Bravo-Marquez, 2017) features a curated selection of tweets, with each tweet labelled under one of four emotion categories. This dataset is integral for understanding emotional expressions in the concise and often nuanced medium of social media, particularly in the realm of Twitter.

Apart from emotion classification, sarcasm classification is also a challenging task for affective analysis in NLP, which features the incongruity between the literal and implied meanings (Riloff et al., 2013). That is, the crux of identifying sarcastic texts lies in the incongruity between the seemingly positive expression and the underlying negative situation. This incongruity is key to discerning sarcastic intent (Riloff et al., 2013). For instance, consider a sarcastic remark such as *"Great, another rainy day – just what I needed!"* While the word "Great" might initially suggest a positive sentiment, the overall context reveals a negative situation, reflecting displeasure towards the continual rain. To test whether our proposed SensoryT5 model can work effectively for other affective analysis tasks, three sarcasm classification datasets are also selected.

- **SemEval 2018** (Van Hee, Lefever, & Hoste, 2018) features a collection of English tweets, curated for Task 3 and Subtask A of the SemEval 2018. The dataset encompasses an array of tweets, each annotated for irony detection. This dataset is significant for its focus on the nuanced task of identifying irony in brief and context-dependent social media texts.
- **Ghosh** (Ghosh & Veale, 2016) is a Twitter-based dataset, leveraging hashtags for automatic sample annotation. It is particularly notable for its use in sarcasm classification, employing advanced neural network techniques to identify sarcasm in social media discourse, making it a valuable resource for studies in computational linguistics and sentiment analysis.
- **IAC-V2** (Abbott, Ecker, Anand, & Walker, 2016) is sourced from online political debates, distinguished by its longer and more structured sample texts compared to other datasets. This dataset is part of the Internet Argument Corpus 2.0, aimed at facilitating research in dialogic social media and the dynamic of online political discussions.

To further evaluate the versatility of our SensoryT5 model, we incorporated one dataset for subjectivity analysis and one dataset for opinion classification. These tasks complement the primary task of emotion classification by providing additional insights.

- **SUBJ** (Pang & Lee, 2004) is a dataset specifically designed for subjectivity analysis, which is closely related to emotion classification. This dataset consists of movie review sentences annotated as either subjective or objective. Subjective sentences express personal opinions, feelings, or evaluations, while objective sentences describe factual information without personal sentiment. Understanding subjectivity is crucial for accurately interpreting emotional content, making this dataset a valuable resource for enhancing emotion classification models by providing a deeper understanding of subjective expressions.
- **PC (Comparative Sentences)** (Ganapathibhotla & Liu, 2008) is a dataset designed for binary classification of opinions in comparative sentences. Each sentence in this dataset is annotated as expressing either a positive or negative opinion about the entities being compared. This task is essential for understanding nuanced opinions in text, particularly when preferences between entities are expressed. By analysing comparative sentences, we can gain a more comprehensive view of how opinions and emotions are conveyed, which complements the primary emotion classification task.

Table 2 presents a summary of the key statistics for these selected datasets. Our evaluation utilizes two widely recognized performance metrics: accuracy and the F1 score.

4.2. Sensory knowledge

Before conducting the emotion analysis experiments, we conducted a preliminary analysis of the sensory lexicon from the perspective of sensory perception value distribution. Fig. 2 displays histograms of the six sensory measures across all words in the Lancaster norms. Notably, the distributions of these perceptual measures are unbalanced. Gustatory and olfactory measures predominantly demonstrate a right-skewed distribution,[1] with most values ranging between 0 and 1. This suggests that these two

[1] A right-skewed distribution, also known as a positive skew, is characterized by a tail that extends more significantly to the right, indicating that a majority of data points are concentrated on the left of the peak. This results in the mean being greater than the median, and the median being greater than the mode (Mean > Median > Mode).

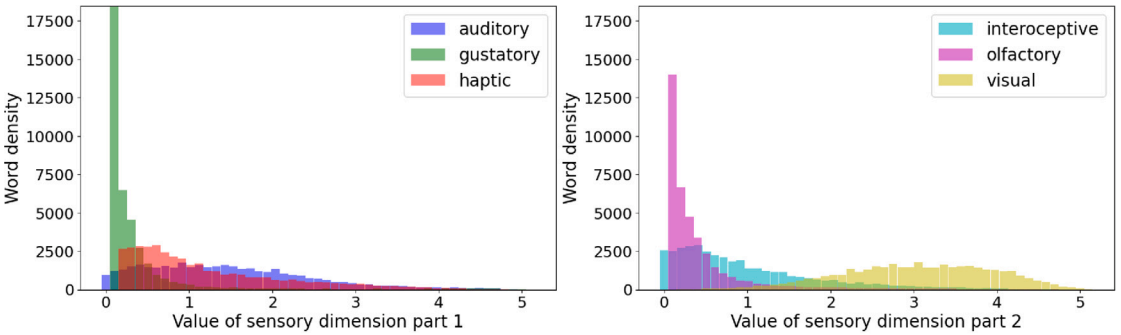

Fig. 2. The distribution of six sensory values over words. The X-axis shows the value in a sensory dimension, and the y-axis displays the word density.

Table 3
Comparison of the accuracy of predictions between T5 embedding and GloVe techniques on different sensory dimensions, as measured by RMSE values. Lower scores indicate higher accuracy in the prediction of sensory values. The values in parentheses following GloVe and Word2Vec indicate the vocabulary size (M) and the number of dimensions (d). GloVe (50d) and GloVe (200d) have a vocabulary size of 0.4M.

Sensory name	T5 embedding	GloVe (50d)	GloVe (200d)	GloVe (2.2M, 300d)	Word2Vec (3M, 300d)
Haptic	0.893	0.764	0.698	**0.640**	0.699
Gustatory	0.632	0.546	0.534	**0.423**	0.464
Olfactory	0.572	0.518	0.501	**0.446**	0.659
Visual	0.842	0.711	0.743	**0.605**	0.621
Auditory	0.949	0.859	0.803	0.703	**0.463**
Interoceptive	0.831	0.755	**0.662**	0.692	0.676
Total	0.798	0.703	0.665	**0.595**	0.605

sensory perceptions are less frequently represented in the textual context. Thus, it might be challenging to represent gustatory and olfactory perceptions from text.

In contrast, auditory and visual measures show a relatively uniform distribution. The auditory measure is evenly distributed between 0 and 2.5, while the visual measure ranges between 2 and 4.5. These distributions indicate a higher sensitivity of auditory and visual knowledge to textual information, which suggests that auditory and visual senses may play a significant role within sensory models. Lastly, haptic and interoceptive measures exhibit similar trends, declining from about 2500 to 0 as the values increase from 0 to 5. These declines in the presence of haptic and interoceptive knowledge across the general textual context might suggest that they are less informative sensory dimensions in the majority of cases.

As discussed in Section 3.1, the Lancaster norms are subject to the size limitation, resulting in a significant number of out-of-vocabulary words whose sensory values are unavailable. To address this challenge, we adopted the method proposed by Li et al. (2017) for predicting sensory values of unknown words through embedding techniques. In our experiments, we utilized the T5 embedding, GloVe embeddings (Pennington, Socher, & Manning, 2014) with three different configurations, and Word2Vec (Mikolov, Chen, Corrado, & Dean, 2013) embedding for this prediction task. We conducted extensive experiments with various versions of GloVe and Word2Vec embeddings to examine the prediction results. Our experiments found that the prediction results are highly correlated with the scale and quality of the pre-trained word embeddings. Here, we present four representative versions: three GloVe versions with significant differences in scale and quality, and the most famous and largest Word2Vec Google News version.

To assess the performance of our predictions, we randomly selected 10% of the Lancaster norms as a validation set and applied the Root Mean Square Error (RMSE) as the evaluation metric. The results of the prediction task are presented in Table 3. Overall, the GloVe embedding with 2.2 million vocabulary size and 300 dimensions outperforms the other embeddings, including the T5 embedding, the smaller versions of GloVe, and the Word2Vec embedding. Following data processing and removal of unusable entries, the size of our sensory vocabulary has expanded to 2,088,280.[2]

For validating our augmentation, we evaluated the coverage rates of sensory word vectors before and after augmentation across the datasets we employed, with the details in Table 4. As evident from the augmentation results, the word coverage has been expanded significantly in comparison to the original data across all the datasets. This would provide an enhanced impact of integrating sensory information into the model on the results.

[2] The entire dataset of the sensory vocabulary can be accessed at: https://osf.io/w8yez/?view_only=0e807dfaa5e6433184e452bfebabd01b.

Table 4
Word coverage of Lancaster norms before and after expansion using regression prediction.

	Datasets	Lancaster %	Expand-Lancaster %
Emotion	GoEmotions	46.85	86.81
	EmD	58.23	94.18
	ISEAR	54.62	93.01
	EmoInt	29.65	65.60
Sarcasm	SemEval 2018	25.26	64.07
	Ghosh	33.27	78.25
	IAC-V2	66.66	92.59
Subjectivity	SUBJ	51.64	93.02
Opinion	PC	52.12	90.96

4.3. Selected baselines

Regarding the selection of baselines, we referred to recent work for emotion classification, sarcasm classification, and related tasks, such as Boonyarat, Liew, and Chang (2024), He et al. (2024), Myint, Lo, and Zhang (2024), Qin and Zhang (2024) and Wan, Wu, Yeo, and Li (2024). Based on these studies, we have selected the following models for comparison.

- **BERT** (Devlin, Chang, Lee, & Toutanova, 2019): A Pretrained Language Model (PLM) that revolutionized text classification by processing text inputs in the [CLS] text [SEP] format. Its deep semantic understanding makes it highly effective for emotion and sarcasm classification tasks.
- **RoBERTa** (Liu et al., 2019): As an enhanced version of BERT, RoBERTa refines training processes and hyperparameters to significantly improve performance in NLP tasks.
- **XLNet** (Yang et al., 2019): This model extends BERT's capabilities by learning bidirectional contexts and using an autoregressive formulation, overcoming some of BERT's limitations.
- **T5** (Raffel et al., 2020): The Text-to-Text Transfer Transformer (T5) adapts all NLP tasks into a text-to-text format, showcasing the versatility and strong performance across various tasks.
- **LLaMA2** (Touvron et al., 2023): LLaMA2 is a state-of-the-art language model that enhances architectural and training processes, achieving impressive performance across various natural language processing tasks. Its extensive pretraining on large datasets contributes to its effectiveness and versatility.
- **LLaMA3** (Meta LLaMA Team, 2024): LLaMA3 builds upon the capabilities of LLaMA2 by introducing deeper semantic understanding and more efficient training methods. It stands as one of the most advanced models for numerous NLP tasks, showcasing significant improvements in performance and handling complex language phenomena.
- **Label-Aware Contrastive Loss** (Suresh & Ong, 2021): This approach assigns different weights to negative samples in classification tasks, enhancing the model's ability to understand fine nuances in text. In emotion classification, it revolutionizes sensitivity and discernment in complex scenarios, particularly adapted to distinguishing closely related emotional states.
- **HypEmo Framework** (Chen, Hung, Hsu, & Ku, 2023): This approach utilizes hyperbolic space for label embedding, offering effective differentiation of subtle label nuances, especially in hierarchical classification tasks. In emotion classification, this framework excels in disentangling subtle emotional nuances, establishing a refined understanding of emotional labels in complex datasets.
- **SarDeCK** (Li, Pan, Lin, Fu and Wang, 2021): This framework integrates commonsense knowledge into sarcasm classification, leveraging contextual and world knowledge for more accurate identification of sarcastic content. This approach marks a significant leap in understanding sarcasm, enabling nuanced and accurate detection.
- **SD-APRR** (Min et al., 2023): This model represents a novel approach in sarcasm classification, augmenting the model's capability with potential results and reactions to emulate a human-like understanding of sarcasm. This innovative method enhances the model's ability to discern sarcasm in complex and varied contexts.
- **BCL (BERT-Base + CLR + LSTM)** (Nandi, Maiya, Kamath, & Shekhar, 2021): This approach integrates BERT-Base for superior text representation with LSTM for enhanced sequence modelling. It achieves state-of-the-art results in subjectivity analysis by fine-tuning the BERT Language Model, significantly improving the classification of subjective versus objective contents. BCL provides a computationally efficient and highly effective solution for nuanced text analysis.
- **DualCL** (Chen, Zhang, Zheng, & Mao, 2022): This framework introduces a dual contrastive learning approach, simultaneously learning the features of input samples and classifier parameters. It enhances classification accuracy by leveraging contrastive learning between input and augmented samples. DualCL improves the discernment of nuanced content, offering robust performance across various text classification tasks, including subjectivity and opinion mining.

4.4. Experiment settings and implementation details

During our experiments, we utilized two NVIDIA A6000 GPUs, each equipped with 48 GB of memory, to train the models. The A6000 GPUs are known for their high performance and large memory capacity, making them well-suited for handling the

Table 5
Configuration and hyperparameters for LLaMA2 and LLaMA3.

Hyper-parameter	Value
BATCH_SIZE	8
ACCUMULATION_STEPS	8
lora_config.r	8
lora_config.lora_alpha	32
lora_config.target_modules	[“q_proj”, “v_proj”]
lora_config.lora_dropout	0.1
lora_config.bias	“none”
Optimizer	AdamW
Learning_rate	2e−5
Scaler	torch.cuda.amp.GradScaler

Table 6
Results of the SensoryT5 model in comparison to the baselines across four emotion classification datasets. The best performances are highlighted in bold and the second-best in underline.

	GoEmotions		EmD		ISEAR		EmoInt	
	ACC	F1	ACC	F1	ACC	F1	ACC	F1
$BERT_{large}$	0.642	0.637	0.588	0.582	0.705	0.700	0.848	0.848
$RoBERTa_{large}$	0.652	0.644	0.596	0.590	0.723	0.720	0.865	0.865
$XLNet_{large}$	0.641	0.568	0.599	0.592	0.711	0.711	0.845	0.845
$T5_{large}$	0.661	0.657	0.609	0.604	0.717	0.717	0.863	0.863
$LLaMA2_{7B}$	0.633	0.629	0.585	0.579	0.680	0.686	0.723	0.817
$LLaMA3_{8B}$	0.650	0.645	0.599	0.600	0.707	0.707	0.760	0.844
LCL	0.655	0.648	0.601	0.591	0.724	**0.724**	0.866	0.866
HypEmo	0.654	0.663	0.596	0.610	0.707	0.712	0.846	0.846
SensoryT5	**0.674**	**0.670**	**0.618**	**0.615**	**0.727**	**0.724**	**0.876**	**0.876**

computational demands of our models. The experiments were conducted on a Linux-based system with CUDA version 11.4, ensuring compatibility with the latest deep learning frameworks and libraries. We used LoRA (Hu et al., 2021) for fine-tuning the LLMs. The specific parameters and configurations for LLaMA2 and LLaMA3 are detailed in Table 5. For the other models, we applied the Adam optimizer in Euclidean space, characterized by its efficiency in navigating the parameter space and adjusting weights to minimize loss. The hyperparameters used for these models are a learning rate of $2e{-}5$ and a batch size of 32.

5. Results and discussion

5.1. Results of experiments

The results of our experiments are shown in Tables 6, 7, and 8. The four tasks are discussed separately in the following.

Emotion classification: In terms of the emotion classification task, SensoryT5 demonstrates a clear improvement in performance over PLMs. When compared with T5 which achieves the best performances among PLMs in the datasets of GoEmotions and EmD, SensoryT5 shows an increase of performances in both accuracy and the F1 score: ACC by 1.3% and F1 by 1.3% for GoEmotions; and ACC by 0.9% and F1 by 1.1% for EmD. These improvements are also observed in the datasets of ISEAR and EmoInt, where SensoryT5 also outperforms RoBERTa, which gains the best performances among PLMs in the datasets of ISEAR and EmoInt, in both accuracy and the F1 score: ACC by 0.4% and F1 by 0.4% for ISEAR; and ACC by 1.1% and F1 by 1.1% for EmoInt.

As shown in Zhang et al. (2023, 2024), the performance of LLMs in emotion classification tasks generally lags behind that of PLMs. Although LLaMA3, as a newly released large model, demonstrates strong capabilities and outperforms many PLMs, its performance still falls short of T5. This trend is also observed with LLaMA2, which does not achieve the same level of accuracy and F1 scores as the top-performing PLMs and SensoryT5.

Further comparisons to the SOTA models such as LCL and HypEmo for emotion classification can also show the superior performance of our proposed SensoryT5 model. While LCL and HypEmo show commendable results, SensoryT5 surpasses LCL consistently across the four datasets: ACC by 1.9% and F1 by 2.2% for GoEmotions; ACC by 1.7% and F1 by 2.4% for EmD; ACC by 0.3% for ISEAR; and ACC by 1.0% and F1 by 1.0% for EmoInt. The improvements of the performances of SensoryT5 over HypEmo are in: ACC by 2.0% and F1 by 0.7% for GoEmotions; ACC by 2.2% and F1 by 0.5% for EmD; ACC by 2.0% and F1 by 1.2% for ISEAR; and ACC by 3.0% and F1 by 3.0% for EmoInt. It is also noteworthy that SensoryT5 does not employ the data augmentation techniques that LCL utilizes.

Sarcasm classification: Turning to sarcasm classification, SensoryT5 also outperforms the PLMs. As shown in Table 7, T5 gains the best performances for the sarcasm classification task across the three selected datasets. However, compared to T5, our proposed SensoryT5 model achieves better accuracy and the F1 score on the datasets: in ACC by 1.2% and F1 by 1.1% for SemEval 2018; ACC by 0.6% and F1 by 0.5% for Ghosh; and ACC by 0.8% and F1 by 0.8% for IAC-V2. Notably, on the Ghosh dataset, LLaMA3

Table 7
Results of the SensoryT5 model in comparison to the baselines across three sarcasm classification datasets. The best performances are highlighted in bold and the second-best in underline.

	SemEval 2018		Ghosh		IAC-V2	
	ACC	F1	ACC	F1	ACC	F1
$BERT_{large}$	0.708	0.706	0.841	0.842	0.791	0.791
$RoBERTa_{large}$	0.731	0.712	0.846	0.848	0.815	0.815
$XLNet_{large}$	0.750	0.727	0.835	0.828	0.815	0.813
$T5_{large}$	0.765	0.768	0.855	0.856	0.822	0.822
$LLaMA2_{7B}$	0.739	0.715	0.802	0.871	0.785	0.789
$LLaMA3_{8B}$	0.760	0.755	0.827	**0.886**	0.806	0.818
SarDeCK	0.717	0.702	0.834	0.830	0.775	0.775
SD-APRR	0.722	0.707	0.826	0.823	0.788	0.788
SensoryT5	**0.777**	**0.779**	**0.861**	0.861	**0.830**	**0.830**

Table 8
Results of the SensoryT5 model in comparison to the baselines across subjectivity analysis dataset and opinion classification dataset. The best performances are highlighted in bold and the second-best in underline.

	SUBJ		PC	
	ACC	F1	ACC	F1
$BERT_{large}$	0.953	0.953	0.942	0.942
$RoBERTa_{large}$	0.961	0.961	0.952	0.952
$XLNet_{large}$	0.956	0.956	0.946	0.946
$T5_{large}$	0.968	0.968	0.954	0.954
$LLaMA2_{7B}$	0.966	0.966	0.952	0.952
$LLaMA3_{8B}$	0.971	0.971	0.954	0.954
BCL	0.973	0.973	NA	NA
DualCL	0.973	NA	0.956	NA
SensoryT5	**0.979**	**0.979**	**0.960**	**0.960**

and LLaMA2 achieve the highest and second-highest F1 scores, with 0.886 and 0.871 respectively, indicating that large language models may have certain potential in sarcasm tasks. However, SensoryT5 still records the highest values for accuracy and F1 score across all datasets, with T5 following closely behind. When compared to specialized sarcasm classification models such as SarDeCK and SD-APRR, SensoryT5 continues to demonstrate its robustness. It records higher accuracy and the F1 score on all the selected sarcasm classification datasets, affirming its capacity to understand and classify sarcastic nuances effectively.

Subjectivity analysis and opinion classification: Shifting focus to the subjectivity analysis and opinion classification tasks, the experimental results, as shown in Table 8, indicate that these tasks are relatively less challenging compared to emotion classification and sarcasm classification. The specialized SOTA models, BCL and DualCL, designed specifically for subjectivity analysis and text classification, have achieved remarkable results in these tasks, surpassing all PLMs and LLMs. However, SensoryT5 continues to demonstrate its superior performance across both datasets. On the SUBJ dataset for the subjectivity analysis task, SensoryT5 exhibits a notable improvement over the second-best performing model, with increases in ACC and F1 scores by 0.6%. Similarly, on the opinion classification dataset, PC, SensoryT5 outperforms the second-best model, achieving an increase in ACC by 0.4% and F1 by 0.6%. Although these numerical improvements may appear modest, they are significant given the already high performance levels in these tasks and the fact that the SOTA models are specifically designed for these purposes. Therefore, the enhancements achieved by SensoryT5 in these tasks are meaningful and noteworthy.

In summary, across emotion classification, sarcasm classification, subjectivity analysis, and opinion classification tasks, SensoryT5 not only outperforms the PLMs and LLMs but also surpasses the current SOTA models. These achievements are particularly significant, as they are accomplished without the need for additional data, thereby setting a new standard in these complex NLP tasks.

5.2. Ablation studies

5.2.1. Evaluating different sensory information integration strategies

To identify the most effective strategy for integrating sensory information into the T5 model, we conducted experiments comparing different methods of sensory integration. These experiments aimed to determine the optimal approach for enhancing the model's performance in emotion classification tasks. The experiments were carried out on two datasets: GoEmotions and EmD. The evaluation involved the following three configurations:

T5 (None): The baseline model without any sensory information, representing the standard approach in fine-grained emotion classification tasks.

Table 9
The performances of T5 (None), T5-SensoryConcat, and SensoryT5 across two emotion classification datasets, evaluated using accuracy as the metric.

	GoEmotions		EmD	
	ACC	F1	ACC	F1
$T5_{large}$	0.661	0.657	0.609	0.604
T5-SensoryConcat	0.659	0.647	0.608	0.597
SensoryT5	**0.674**	**0.670**	**0.618**	**0.615**

T5-SensoryConcat: In this configuration, we experimented with various concatenation methods. For instance, we tried directly integrating the 6 dimensional sensory information into the T5 embedding to form a 1030 dimensional vector, which was then reduced to 1024 dimensions. Another approach expanded the 6 dimensional sensory information to 1024 dimensions, concatenated it with the T5 embedding to create a 2048 dimensional vector, and subsequently reduced it to 1024 dimensions. However, these methods led to significant information loss, disrupting the original T5 embedding structure and resulting in suboptimal performance. Finally, the best performing method involved replacing the last six dimensions of the T5 embedding with the 6 dimensional sensory information.

SensoryT5: Our proposed model, which integrates cognitively-grounded and linguistically-encoded sensory knowledge into the T5 architecture through an additional attention mechanism specifically designed to fuse sensory information with contextual embeddings.

As shown in Table 9, the performance of the feature concatenation method for embedding sensory information into the T5 embedding was unsatisfactory, even yields lower results than the T5 (None) model. This outcome was anticipated, as the composition and structural content of the T5 embedding and sensory information are fundamentally different. While traditional NLP methods have seen success by concatenating external data features to word embeddings in certain tasks, this approach is not effective in transformer models. Transformer models are pre-trained on vast corpora, encapsulating rich semantic and contextual information. Introducing external data directly does not enhance the model's performance and may even introduce noise. The T5 embeddings have been finely tuned during pre-training, possessing specific structures and distributions. Unnecessary expansion and scaling of these embeddings could disrupt this structure, weakening the semantic representations and finally reducing model performance.

In contrast, attention-based models have successfully incorporated external data resources for sentiment classification in numerous studies. For example, Long et al. (2019) proposed an attention mechanism trained with cognitively-grounded eye-tracking data, significantly improving sentiment analysis performance. Similarly, Ayetiran (2022) proposed an attention-based model that enhances aspect-based sentiment classification by integrating external document-level information, thereby improving the model's ability to capture contextual and semantic features. Therefore, we chose to integrate sensory information through a custom sensory attention structure, which preserves the original T5 contextual embeddings. The results indicate that our sensory information integration strategy is effective.

5.2.2. Evaluating the impact of sensory knowledge

To understand the contribution of different components within the SensoryT5 model, we conducted ablation studies, which are crucial in assessing the impact of our novel sensory integration. These studies were carried out on the four datasets for the fine-grained emotion classification task: GoEmotions, EmD, ISEAR, and EmoInt. The ablation tests were structured based on the following three primary configurations.

T5 (None): The baseline model without any sensory information, representing the standard approach in fine-grained emotion classification tasks.

Random SensoryT5: A variant of our model where the sensory values were substituted with random numbers ranging from 0 to 5, maintaining the same distribution of sensory scores but eliminating their meaningful associations with the sensory vocabulary data.

SensoryT5: Our proposed model with the cognitively-grounded and linguistically-encoded sensory knowledge infused.

The results are shown in Fig. 3. Although the SensoryT5 model exhibits the best performances in terms of accuracy across all the selected datasets, the Random SensoryT5 yields lower results than the T5 (None) model. The decrease in the performance is more evident on the more complex datasets, i.e., GoEmotions and EmD. This tendency underscores the importance of meaningful sensory integration. Thus, we can argue that it is not merely the presence of additional numerical data that has enhanced the SensoryT5 model's performance, but rather the cognitively-grounded and linguistically-encoded sensory knowledge (Lynott et al., 2020) has contributed to the emotion classification. On the other hand, the fact that the Random SensoryT5 model works less effectively compared to the T5 (None) model indicates that arbitrarily added sensory information could introduce noise into the model, disrupting its ability to correctly interpret and classify emotional content. This finding also affirms that strategic integration of sensory data is crucial, and haphazard integration could be counterproductive.

In summary, these ablation studies have confirmed the value of our sensory information layer, as evidenced by the performance drop when this layer is removed or randomized. This reinforces our assertion that the SensoryT5's strength lies in its ability to simulate a more human-like understanding, resonating with how humans perceive emotions through a sensory lens.

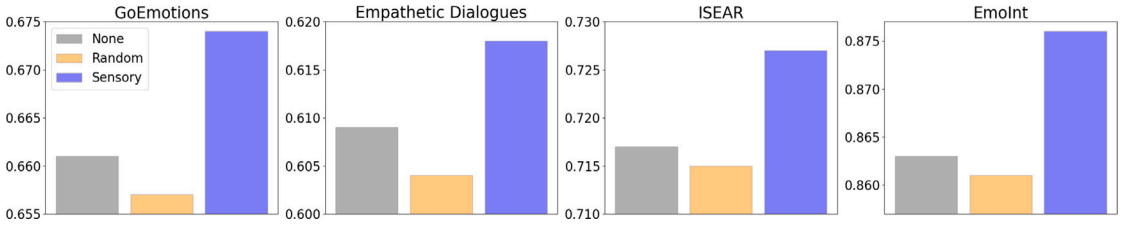

Fig. 3. The performances of T5 (None), Random SensoryT5 (with sensory values randomly assigned), and SensoryT5 across four emotion classification datasets, evaluated using accuracy as the metric.

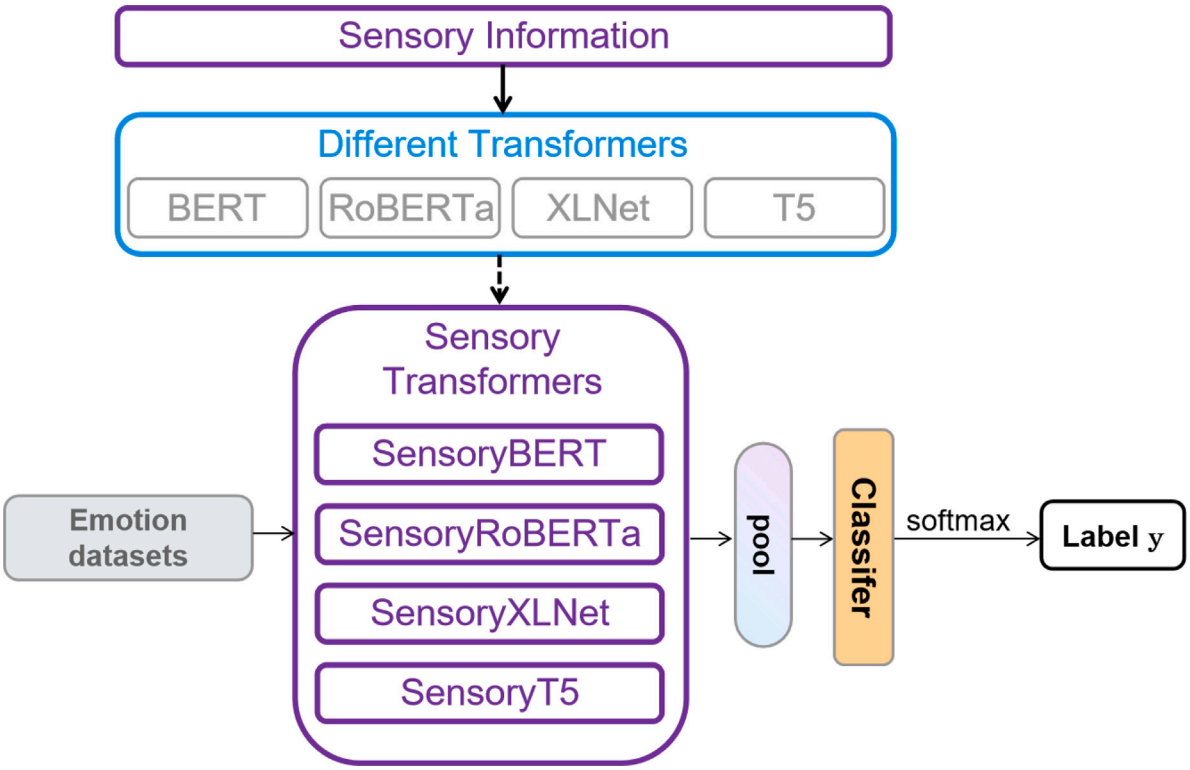

Fig. 4. Infusing sensory knowledge into different transformer-based models. In SensoryXLNet, both K (key) and V (value) in the sensory attention layer are constituted by the output of the final hidden layer of XLNet.

5.2.3. Impact of sensory knowledge on different foundation models

To further explore the versatility and efficacy of the sensory knowledge in NLP, we extended our experiments by infusing the cognitively-motivated and linguistically-encoded sensory data into other foundational transformer models, including BERT, RoBERTa, and XLNet, as shown in Fig. 4. The aim is to observe the impact of sensory knowledge on these well-established models for emotion classification.

The results of these experiments are summarized in Table 10. Upon integrating the sensory knowledge with BERT, a noticeable improvement is observed in both accuracy and the F1 score across the GoEmotions and EmD datasets. Specifically, SensoryBERT outperforms the original BERT model, showing an increase in accuracy by 1.3% and the F1 score by 1.4% for the dataset of GoEmotions. A slight enhancement has also been achieved for integrating the sensory knowledge with BERT over the original BERT model, in accuracy by 0.8% and the F1 score by 0.7% for the dataset of EmD. Similarly, the integrations of sensory knowledge with RoBERTa (SensoryRoBERTa) and XLNet (SensoryXLNet) have also yielded positive results. For sensoryRoBERTa, there are improvements in accuracy by 0.8% and the F1 score by 1.4% for GoEmotions as well as in accuracy by 0.6% and the F1 score by 0.5% for EmD. In the case of sensoryXLNet, the improvements are in: ACC by 0.9% and F1 by 0.7% for GoEmotions; and ACC by 0.6% and F1 by 0.7% for EmD.

Most notably, the sensory enhancement with T5 (SensoryT5) demonstrates the most significant improvements. SensoryT5 achieves an increase of 0.9% in accuracy and 1.1% in the F1 score for EmD as well as 1.3% in both accuracy and the F1 score for GoEmotions, outperforming the already robust results of the original T5 model.

Table 10
Impact of sensory knowledge on different foundation models.

	GoEmotions		EmD	
	ACC	F1	ACC	F1
$BERT_{large}$	0.642	0.637	0.588	0.582
SensoryBERT	**0.655**	**0.651**	**0.596**	**0.589**
$RoBERTa_{large}$	0.652	0.644	0.596	0.590
SensoryRoBERTa	**0.660**	**0.658**	**0.602**	**0.595**
$XLNet_{large}$	0.641	0.568	0.599	0.592
SensoryXLNet	**0.650**	**0.577**	**0.605**	**0.597**
$T5_{large}$	0.661	0.657	0.609	0.604
SensoryT5	**0.674**	**0.670**	**0.618**	**0.615**

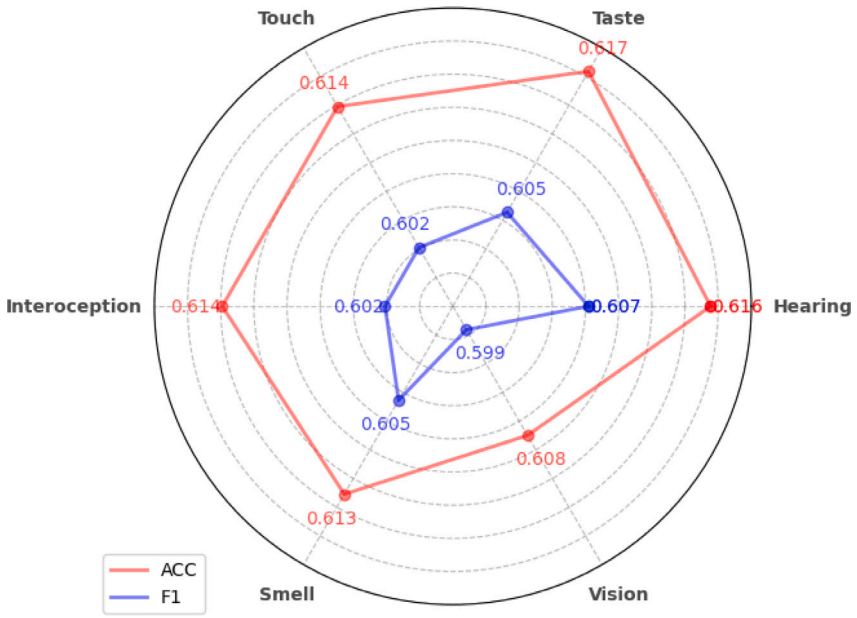

Fig. 5. The radar chart depicts SensoryT5's performance on the Empathetic Dialogue dataset with the removal of one single sensory dimension. Each axis corresponds to the model's performance without the named sensory input. The chart's vertices represent the model's accuracy and F1 score after excluding each particular sensory dimension. The closer a data point is to the outer edge of the chart, the higher the performance achieves.

These results collectively demonstrate the value of the infused sensory knowledge across different foundational models. The consistent improvements indicate that the sensory method is not only effective but also adaptable to various transformer architectures, amplifying their capabilities in emotion classification tasks. Furthermore, as the sensory knowledge represents an important kind of cognition-anchored resource, the results show the potential and the important value of neuro-cognitive data and computational approaches synergized in NLP studies.

5.2.4. Identifying critical sensory dimensions in SensoryT5

To further understand sensory experiences for emotion classification, we conducted ablation studies on sensory dimensions employed in the SensoryT5 model, including touch, taste, smell, vision, hearing, and interoception. The results of removing one single sensory dimension and removing two combined sensory dimensions are presented in Figs. 5 and 6 respectively.

In terms of the removal of one single sensory dimension, the performance of SensoryT5 drops most drastically in both accuracy and the F1 score on the Empathetic Dialogue dataset when vision is removed. This indicates that vision is the most critical sensory dimension for the performance of SensoryT5. This finding is consistent with a body of neuroscience research that has established a strong link between visual processing and emotional understanding. For example, Nanda, Zhu, and Jansen (2012) have demonstrated that visual stimuli induce strong and distinct patterns of brain activation, particularly in the amygdala, which is heavily involved in the processing of emotions such as fear, anxiety, and pain. Additionally, the visual cortex is known to be activated during the processing of emotional images, further supporting the notion that emotional processing begins with vision (Lang et al., 1998). Furthermore, recall that visual knowledge is the most widely represented in the text (see Fig. 2), which might be one of the reasons that vision contributes to the SensoryT5 model most greatly. In other words, other sensory dimensions such as taste and smell, while

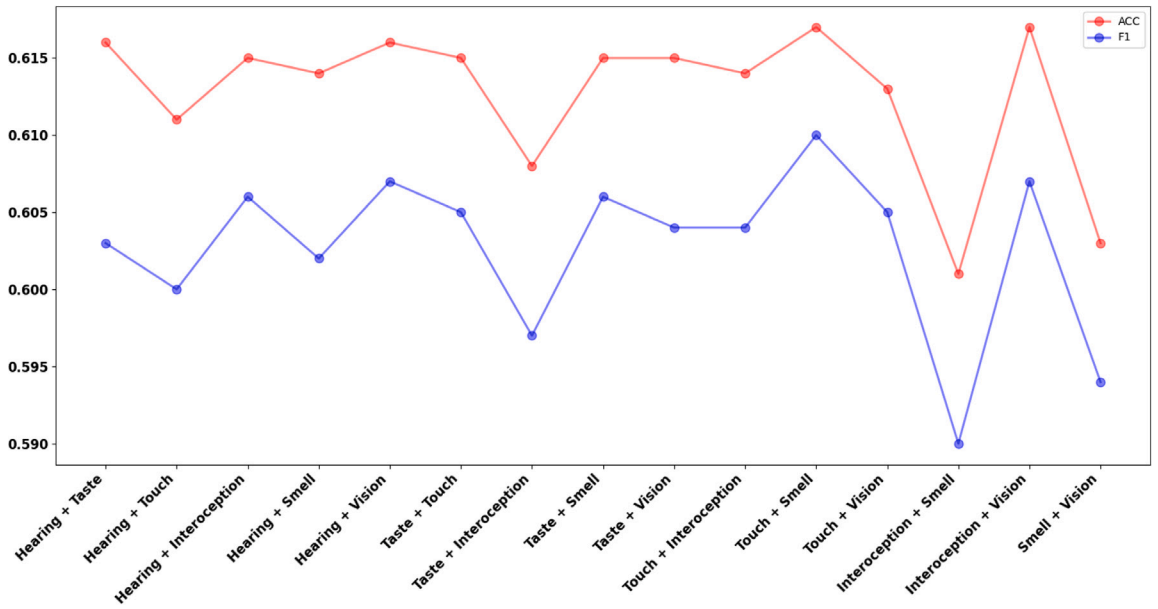

Fig. 6. This line graph presents the performance of SensoryT5 on the Empathetic Dialogue dataset when pairs of sensory dimensions are simultaneously ablated. Each point on the graph represents the model's accuracy (ACC) and F1 score after the removal of two combined sensory inputs.

important and also closely related to emotions (Kadohisa, 2013; Mastinu, Melis, Yousaf, Barbarossa, & Tepper, 2023; Winter, 2016), showing a smaller impact on the model's performance, may be due to their less-frequently representations in the textual context.

When pairs of sensory dimensions are simultaneously ablated, the performance of the SensoryT5 model drops most drastically without interoception and smell (see Fig. 6). This echoes that emotion is a kind of interoception (Connell et al., 2018; Lynott et al., 2020) and that emotion is intimately connected to smell (Kadohisa, 2013; Mastinu et al., 2023; Winter, 2016). However, the pair of vision and smell also contributes to the performance of SensoryT5 greatly, with a slightly smaller impact than the pair of interoception and smell. This pattern again shows that visual knowledge is fundamental to emotion categorization and processing.

To summarize, our ablation studies find that vision shows the most important impact on the performance of the SensoryT5 model among the sensory dimensions. This might suggest that when SensoryT5 loses the visual input, it might be akin to the brain receiving less information to initiate and process emotional responses. Thus, the SensoryT5 model might not only rely heavily on visual knowledge but also indicate that the model finds it more challenging to integrate sensory inputs into a coherent emotional understanding without the grounding context that visual information provides.

5.3. Case study

Case studies are presented in Fig. 7 using four sentences from the Empathetic Dialogues, GoEmotions, ISEAR, and EmoInt datasets: *"I get so mad when I see or hear about kids getting bullied..."* (Empathetic Dialogues) *"Upon reading I find this to be fake news and I'm disgusted..."* (GoEmotions) *"There is a certain person, whom I only have seen. He makes me cringe, feel disgust."* (ISEAR) *"That's so pretty! I love the sky in the background and the purple highlights with the dull colors is great."* (EmoInt) The SensoryT5 heatmaps illustrate the aggregate attention for each token in the sensory layer, while the T5 section compiles and averages attention weights across all encoder layers to reveal the model's overall focus. The SensoryT5 model shows intensified attention on emotionally significant phrases: *"so mad"* in the first sentence, *"disgusted"* in the second, *"cringe"* in the third, and *"love"* and *"great"* in the fourth. This indicates the model's ability to detect crucial emotional nuances. In contrast, the standard T5's attention is more distributed and less focused on these emotional pivots. These micro-level analyses reveal SensoryT5's superior capability in recognizing emotional cues, substantiating the efficacy of integrating sensory awareness into language models for improved emotion discernment. By selecting diverse sentences from multiple datasets, we demonstrate SensoryT5's robustness and generalizability in identifying and focusing on emotionally charged phrases across different emotional contexts. These micro-level analyses reveal SensoryT5's superior capability in recognizing emotional cues, which substantiate the efficacy of integrating sensory awareness into language models for improved emotion discernment.

5.4. Computational cost analysis

While SensoryT5 introduces significant improvements in emotion classification tasks, it is important to acknowledge the additional computational costs incurred by this model. The primary source of this increased cost stems from the extra sensory

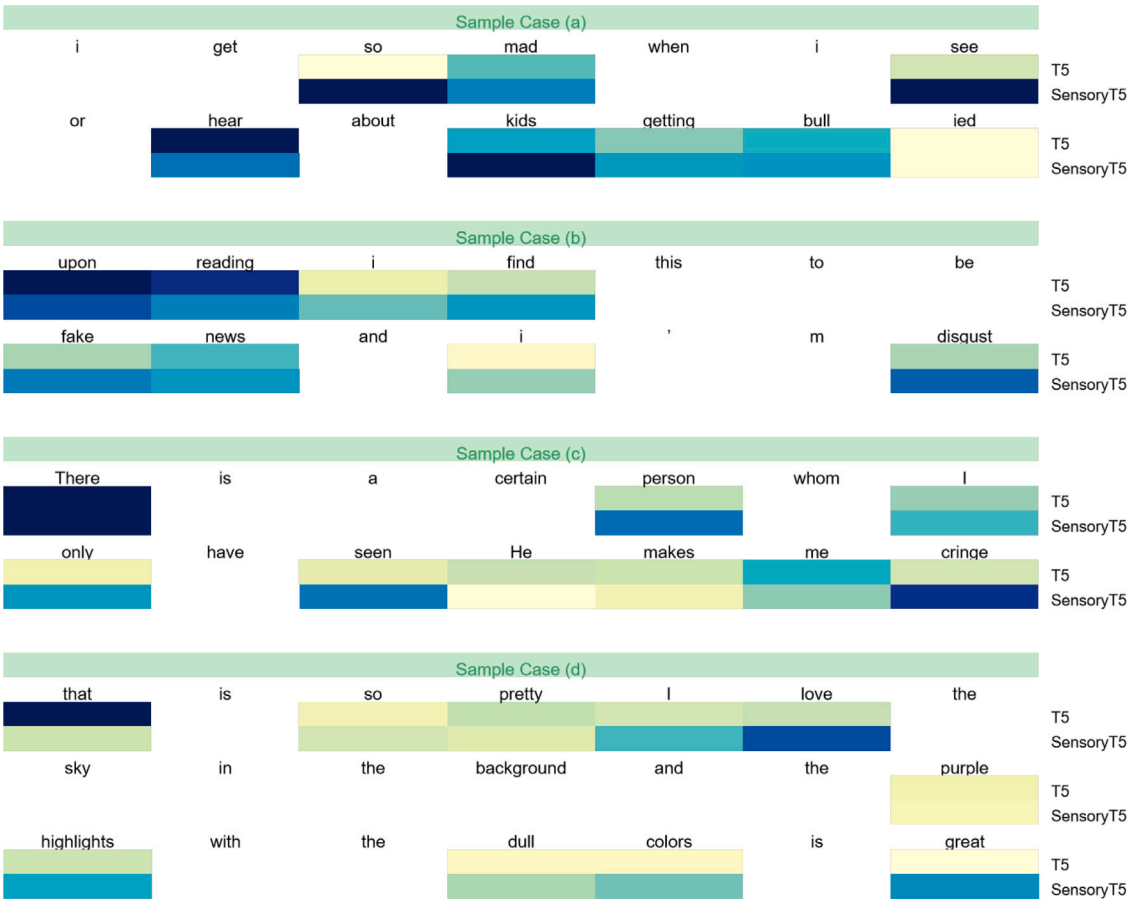

Fig. 7. The heatmap visualizes the heat values of the final sensory layer in SensoryT5 and the encoder layer in T5 for four sentences. Darker colours indicate higher attention weights. These sentences are sourced sequentially from the Empathetic Dialogues, GoEmotions, ISEAR, and EmoInt emotion classification datasets. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

attention layer and the sensory information transformation required for T5 integration. To quantify this, we conducted a comparative analysis of the training time per epoch and the total number of epochs required for convergence between SensoryT5 and the baseline T5 model. Specifically, we used the Opinion Polarity (PC) dataset as a benchmark. On average, each epoch for the T5 model took approximately 3 min and 10 s, while the SensoryT5 model required about 3 min and 30 s per epoch. This represents an increase of approximately 1.11 times in the epoch duration for SensoryT5. Furthermore, SensoryT5 typically requires one additional epoch to achieve optimal results compared to T5. This is attributed to the more complex backpropagation process necessitated by the sensory integration, which in turn extends the time needed for the model to converge.

Despite the marginally higher computational cost, we believe that the enhanced performance and deeper emotional understanding provided by SensoryT5 justify this expense. The trade-off between a slight increase in training time and the substantial gains in classification accuracy and F1 scores demonstrates the cost-effectiveness of SensoryT5 for advanced emotion analysis tasks.

5.5. Summary of experiments

The evaluations and comparative studies demonstrate the superior performance of SensoryT5 over other emotion classification models. When benchmarked against the state-of-the-art methods, SensoryT5 notably surpasses them, establishing a new standard in the field. In addition, our ablation studies consistently attest that the effectiveness of SensoryT5 is attributed to its integration of cognition-grounded sensory knowledge, rather than the mere structural enhancements. This assertion is corroborated by our detailed case studies, which offer a microscopic view into the instances where SensoryT5's unique capabilities are distinctly evident. Furthermore, this study also attests to the visual knowledge being fundamental to emotion categorization and processing, which furthers our understanding of human emotions. Collectively, these findings underscore a breakthrough performance of SensoryT5 in fine-grained emotion classification. The model signifies a successful adaptation within the shift towards incorporating neuro-cognitive data in NLP studies, validating the premise that a deeper convergence between sensory data and language modelling contributes to a more profound understanding of emotion nuances for machines.

6. Conclusion

This study proposes the SensoryT5 model designed for fine-grained emotion classification and sarcasm classification. This framework harnesses sensory knowledge, aiming to boost the prowess of transformers in pinpointing nuanced emotional subtleties. By integrating sensory knowledge into T5 through attention mechanisms, the model concurrently evaluates sensory cues alongside contextual hallmarks. Our comprehensive experiments show that SensoryT5 outperforms the state-of-the-art models, establishing a new standard for tasks in fine-grained emotion classification and sarcasm classification. Moreover, SensoryT5 serves as a conduit between sensory perception and emotional understanding, embodying the recent paradigm shift in NLP studies towards a more neuro-cognitive approach. The model acknowledges and capitalizes on the intrinsic relationship between our sensory experiences and emotional responses, a connection well-documented in neuro-cognitive science but often under-explored in computational fields. By interpreting the sensory lexicon through advanced representation learning, SensoryT5 decodes the implicit emotional undertones conveyed, mirroring the human ability to associate sensory experiences with specific emotional states. In recognizing the entwined nature of cognition, sensation, and emotive expression, SensoryT5 not only contributes to but also encourages the continuation of interdisciplinary research efforts. It stands as a testament to the potential of a more nuanced and integrative approach in NLP, where understanding language transcends the boundaries of words and grammar, delving into the very experiences and perceptions that shape human emotion and cognition.

7. Limitation and future plan

In our work, we have employed GloVe, Word2Vec, and T5 embeddings to predict sensory values for unknown words using a regression method. Based on the findings of Chersoni, Xiang, Lu, and Huang (2020), which demonstrate the feasibility of cross-lingual methods for predicting sensory knowledge, we envision the potential for future applications of our SensoryT5 model in tasks involving other languages. This expansion might significantly enhance the model's versatility and effectiveness in diverse linguistic environments.

CRediT authorship contribution statement

Qingqing Zhao: Writing – review & editing, Writing – original draft, Supervision, Formal analysis, Conceptualization. **Yuhan Xia:** Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation. **Yunfei Long:** Supervision, Project administration, Methodology, Conceptualization. **Ge Xu:** Software, Resources. **Jia Wang:** Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Yunfei Long, Yuhan Xia reports financial support and article publishing charges were provided by University of Essex. Yunfei Long used to act as reviewing member of Information Processing & Management Journal. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work is supported by the Alan Turning Institute/DSO grant: Improving multimodality misinformation detection with affective analysis. Yunfei Long, and Yuhan Xia acknowledge the financial support of the School of Computer science and Electrical Engineering, University of Essex.

References

Abbott, R., Ecker, B., Anand, P., & Walker, M. (2016). Internet argument corpus 2.0: An sql schema for dialogic social media and the corpora to go with it. In *Proceedings of the tenth international conference on language resources and evaluation* (pp. 4445–4452).

Ayetiran, E. F. (2022). Attention-based aspect sentiment classification using enhanced learning through CNN-BiLSTM networks. *Knowledge-Based Systems*, *252*, Article 109409.

Boonyarat, P., Liew, D. J., & Chang, Y.-C. (2024). Leveraging enhanced BERT models for detecting suicidal ideation in thai social media content amidst COVID-19. *Information Processing & Management*, *61*(4), Article 103706.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901.

Chen, M. (2022). Emotion analysis based on deep learning with application to research on development of western culture. *Frontiers in Psychology*, *13*, Article 911686.

Chen, X., Hai, Z., Wang, S., Li, D., Wang, C., & Luan, H. (2021). Metaphor identification: A contextual inconsistency based neural sequence labeling approach. *Neurocomputing*, *428*, 268–279.

Chen, J., Huang, Z., & Xue, Y. (2021). Bilateral-brain-like semantic and syntactic cognitive network for aspect-level sentiment analysis. In *2021 international joint conference on neural networks* (pp. 1–8). IEEE.

Chen, C.-Y., Hung, T.-M., Hsu, Y.-L., & Ku, L.-W. (2023). Label-aware hyperbolic embeddings for fine-grained emotion classification. arXiv preprint arXiv:2306.14822.

Chen, Q., Zhang, R., Zheng, Y., & Mao, Y. (2022). Dual contrastive learning: Text classification via label-aware data augmentation. arXiv preprint arXiv:2201.08702.

Chersoni, E., Xiang, R., Lu, Q., & Huang, C.-R. (2020). Automatic learning of modality exclusivity norms with crosslingual word embeddings. In *Proceedings of the ninth joint conference on lexical and computational semantics* (pp. 32–38).

Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., et al. (2022). Palm: Scaling language modeling with pathways. arXiv preprint arXiv:2204.02311.

Connell, L., Lynott, D., & Banks, B. (2018). Interoception: the forgotten modality in perceptual grounding of abstract and concrete concepts. *Philosophical Transactions of the Royal Society, Series B (Biological Sciences)*, *373*(1752), Article 20170143.

Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., & Ravi, S. (2020). GoEmotions: A dataset of fine-grained emotions. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 4040–4054). Online: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2020.acl-main.372, URL: https://aclanthology.org/2020.acl-main.372.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)* (pp. 4171–4186). Minneapolis, Minnesota: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/N19-1423, URL: https://aclanthology.org/N19-1423.

Duéñez-Guzmán, E. A., Sadedin, S., Wang, J. X., McKee, K. R., & Leibo, J. Z. (2023). A social path to human-like artificial intelligence. *Nature Machine Intelligence*, *5*(11), 1181–1188.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*(3–4), 169–200.

Fainsilber, L., & Ortony, A. (1987). Metaphorical uses of language in the expression of emotions. *Metaphor and Symbol*, *2*(4), 239–250.

Fan, T., Qiu, S., Wang, Z., Zhao, H., Jiang, J., Wang, Y., et al. (2023). A new deep convolutional neural network incorporating attentional mechanisms for ECG emotion recognition. *Computers in Biology and Medicine*, *159*, Article 106938.

Ganapathibhotla, M., & Liu, B. (2008). Mining opinions in comparative sentences. In *Proceedings of the 22nd international conference on computational linguistics* (pp. 241–248).

Ghosh, A., & Veale, T. (2016). Fracking sarcasm using neural network. In *Proceedings of the 7th workshop on computational approaches to subjectivity, sentiment and social media analysis* (pp. 161–169).

Gibbs, R. W., Jr. (2005). *Embodiment and cognitive science*. Cambridge University Press.

Hasim Sak, A. S., & Beaufays, F. (2014). Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition. arXiv preprint arXiv:1402.1128.

He, X., Yu, L., Tian, S., Yang, Q., Long, J., & Wang, B. (2024). VIEMF: Multimodal metaphor detection via visual information enhancement with multimodal fusion. *Information Processing & Management*, *61*(3), Article 103652.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., et al. (2021). Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685.

Kadohisa, M. (2013). Effects of odor on emotion, with implications. *Frontiers in Systems Neuroscience*, *7*, 66.

Khanam, S., Tanweer, S., Khalid, S., & Rosaci, D. (2019). Artificial intelligence surpassing human intelligence: factual or hoax. *The Computer Journal*, *64*(12), 1832–1839.

Khare, S. K., Blanes-Vidal, V., Nadimi, E. S., & Acharya, U. R. (2024). Emotion recognition and artificial intelligence: A systematic review (2014–2023) and research recommendations. *Information Fusion*, Article 102019.

Kim, Y. (2014). Convolutional neural networks for sentence classification. In *Proceedings of the 2014 conference on empirical methods in natural language processing* (pp. 1746–1751). Doha, Qatar: Association for Computational Linguistics, http://dx.doi.org/10.3115/v1/D14-1181, URL: https://aclanthology.org/D14-1181.

Kövecses, Z. (2019). Perception and metaphor. *Perception Metaphors*, *19*(327), 10–1075.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago.

Lang, P. J., Bradley, M. M., Fitzsimmons, J. R., Cuthbert, B. N., Scott, J. D., Moulder, B., et al. (1998). Emotional arousal and activation of the visual cortex: an fMRI analysis. *Psychophysiology*, *35*(2), 199–210.

Lee, S. Y. M. (2018). Figurative language in emotion expressions. In *Chinese lexical semantics: 18th workshop, CLSW 2017, Leshan, China, May 18–20, 2017, revised selected papers 18* (pp. 408–419). Springer.

Li, M., Lu, Q., Long, Y., & Gui, L. (2017). Inferring affective meanings of words from word embedding. *IEEE Transactions on Affective Computing*, *8*(4), 443–456.

Li, J., Pan, H., Lin, Z., Fu, P., & Wang, W. (2021). Sarcasm detection with commonsense knowledge. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *29*, 3192–3201.

Li, Y., Zhang, K., Wang, J., & Gao, X. (2021). A cognitive brain model for multimodal sentiment analysis based on attention neural networks. *Neurocomputing*, *430*, 159–173.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., et al. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.

Long, Y., Xiang, R., Lu, Q., Huang, C.-R., & Li, M. (2019). Improving attention model based on cognition grounded data for sentiment analysis. *IEEE Transactions on Affective Computing*, *12*(4), 900–912.

Lu, Q., Sun, X., Long, Y., Gao, Z., Feng, J., & Sun, T. (2023). Sentiment analysis: Comprehensive reviews, recent advances, and open challenges. *IEEE Transactions on Neural Networks and Learning Systems*.

Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods*, *41*(2), 558–564.

Lynott, D., & Connell, L. (2013). Modality exclusivity norms for 400 nouns: The relationship between perceptual experience and surface word form. *Behavior Research Methods*, *45*, 516–526.

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The lancaster sensorimotor norms: multidimensional measures of perceptual and action strength for 40,000 english words. *Behavior Research Methods*, *52*, 1271–1291.

Mastinu, M., Melis, M., Yousaf, N. Y., Barbarossa, I. T., & Tepper, B. J. (2023). Emotional responses to taste and smell stimuli: Self-reports, physiological measures, and a potential role for individual and genetic factors. *Journal of Food Science*, *88*(S1), A65–A90.

Meta LLaMA Team (2024). Introducing meta llama 3: The most capable openly available LLM to date. https://ai.meta.com/blog/meta-llama-3/.

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, *48*(7), 788–804.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Min, C., Li, X., Yang, L., Wang, Z., Xu, B., & Lin, H. (2023). Just like a human would, direct access to sarcasm augmented with potential result and reaction. In *Proceedings of the 61st annual meeting of the association for computational linguistics (volume 1: long papers)* (pp. 10172–10183).

Mitra, V., Nie, J., & Azemi, E. (2024). Investigating salient representations and label variance in dimensional speech emotion analysis. In *ICASSP 2024-2024 IEEE international conference on acoustics, speech and signal processing* (pp. 11111–11115). IEEE.

Mohammad, S., & Bravo-Marquez, F. (2017). WASSA-2017 shared task on emotion intensity. In *Proceedings of the 8th workshop on computational approaches to subjectivity, sentiment and social media analysis* (pp. 34–49). Copenhagen, Denmark: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/W17-5205, URL: https://aclanthology.org/W17-5205.

Müller, N., Nagels, A., & Kauschke, C. (2021). Metaphorical expressions originating from human senses: Psycholinguistic and affective norms for german metaphors for internal state terms (MIST database). *Behavior Research Methods*, 1–13.

Myint, P. Y. W., Lo, S. L., & Zhang, Y. (2024). Unveiling the dynamics of crisis events: Sentiment and emotion analysis via multi-task learning with attention mechanism and subject-based intent prediction. *Information Processing & Management*, *61*(4), Article 103695.

Nanda, U., Zhu, X., & Jansen, B. (2012). Image and emotion: From outcomes to brain behavior. *HERD: Health Environments Research & Design Journal*, *5*(4), 40–59.

Nandi, R., Maiya, G., Kamath, P., & Shekhar, S. (2021). An empirical evaluation of word embedding models for subjectivity analysis tasks. In *2021 international conference on advances in electrical, computing, communication and sustainable technologies* (pp. 1–5). IEEE.

Nosta, J. (2023). The dawn of sensory AI. URL: https://www.psychologytoday.com/intl/blog/the-digital-self/202309/the-dawn-of-sensory-ai. (Accessed 25 September 2023).

OpenAI (2023). GPT-4 technical report. arXiv:2303.08774.

Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. arXiv preprint cs/0409058.

Peng, L., Zhang, Z., Pang, T., Han, J., Zhao, H., Chen, H., et al. (2024). Customising general large language models for specialised emotion recognition tasks. In *ICASSP 2024-2024 IEEE international conference on acoustics, speech and signal processing* (pp. 11326–11330). IEEE.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing* (pp. 1532–1543).

Přibáň, P., Šmíd, J., Steinberger, J., & Mištera, A. (2024). A comparative study of cross-lingual sentiment analysis. *Expert Systems with Applications*, *247*, Article 123247.

Qin, S., & Zhang, M. (2024). Boosting generalization of fine-tuning BERT for fake news detection. *Information Processing & Management*, *61*(4), Article 103745.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., et al. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, *21*(1), 5485–5551.

Raheel, A., Majid, M., Alnowami, M., & Anwar, S. M. (2020). Physiological sensors based emotion recognition while experiencing tactile enhanced multimedia. *Sensors*, *20*(14), 4037.

Rashkin, H., Smith, E. M., Li, M., & Boureau, Y.-L. (2019). Towards empathetic open-domain conversation models: A new benchmark and dataset. In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 5370–5381). Florence, Italy: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/P19-1534, URL: https://aclanthology.org/P19-1534.

Riloff, E., Qadir, A., Surve, P., De Silva, L., Gilbert, N., & Huang, R. (2013). Sarcasm as contrast between a positive sentiment and negative situation. In *Proceedings of the 2013 conference on empirical methods in natural language processing* (pp. 704–714).

Rodriguez, P., Cucurull, G., Gonzàlez, J., Gonfaus, J. M., Nasrollahi, K., Moeslund, T. B., et al. (2022). Deep pain: Exploiting long short-term memory networks for facial expression classification. *IEEE Transactions on Cybernetics*, *52*(5), 3314–3324.

Romero, O. J., Zimmerman, J., Steinfeld, A., & Tomasic, A. (2023). Synergistic integration of large language models and cognitive architectures for robust ai: An exploratory analysis. arXiv preprint arXiv:2308.09830.

Satpute, A. B., Kang, J., Bickart, K. C., Yardley, H., Wager, T. D., & Barrett, L. F. (2015). Involvement of sensory regions in affective experience: a meta-analysis. *Frontiers in Psychology*, *6*, 1860.

Scherer, K. R., & Wallbott, H. G. (1994). Evidence for universality and cultural variation of differential emotion response patterning. *Journal of Personality and Social Psychology*, *66*(2), 310.

Šimić, G., Tkalčić, M., Vukić, V., Mulc, D., Španić, E., Šagud, M., et al. (2021). Understanding emotions: Origins and roles of the amygdala. *Biomolecules*, *11*(6), 823.

Skaramagkas, V., Ktistakis, E., Manousos, D., Kazantzaki, E., Tachos, N. S., Tripoliti, E., et al. (2023). eSEE-d: Emotional state estimation based on eye-tracking dataset. *Brain Sciences*, *13*(4), 589.

Suresh, V., & Ong, D. (2021). Not all negatives are equal: Label-aware contrastive loss for fine-grained text classification. In *Proceedings of the 2021 conference on empirical methods in natural language processing* (pp. 4381–4394). Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2021.emnlp-main.359, URL: https://aclanthology.org/2021.emnlp-main.359.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., et al. (2023). Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971.

Tuncer, T., Dogan, S., & Acharya, U. R. (2021). Automated accurate speech emotion recognition system using twine shuffle pattern and iterative neighborhood component analysis techniques. *Knowledge-Based Systems*, *211*, Article 106547.

Van Hee, C., Lefever, E., & Hoste, V. (2018). Semeval-2018 task 3: Irony detection in english tweets. In *Proceedings of the 12th international workshop on semantic evaluation* (pp. 39–50).

Wan, M., Su, Q., Ahrens, K., & Huang, C.-R. (2023). Perceptional and actional enrichment for metaphor detection with sensorimotor norms. *Natural Language Engineering*, 1–29.

Wan, B., Wu, P., Yeo, C. K., & Li, G. (2024). Emotion-cognitive reasoning integrated BERT for sentiment analysis of online public opinions on emergencies. *Information Processing & Management*, *61*(2), Article 103609.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*, 625–636.

Wilson, A. D., & Golonka, S. (2013). Embodied cognition is not what you think it is. *Frontiers in Psychology*, *4*, 58.

Winter, B. (2016). Taste and smell words form an affectively loaded and emotionally flexible part of the english lexicon. *Language, Cognition and Neuroscience*, *31*(8), 975–988.

Yamamoto, T. (2008). Central mechanisms of taste: Cognition, emotion and taste-elicited behaviors. *Japanese Dental Science Review*, *44*(2), 91–99.

Yan, X., Zhang, Y., & Zhang, C. (2024). Utilizing cognitive signals generated during human reading to enhance keyphrase extraction from microblogs. *Information Processing & Management*, *61*(2), Article 103614.

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in Neural Information Processing Systems*, *32*.

Zadra, J. R., & Clore, G. L. (2011). Emotion and perception: The role of affective information. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(6), 676–685.

Zhang, W., Deng, Y., Liu, B., Pan, S. J., & Bing, L. (2023). Sentiment analysis in the era of large language models: A reality check. arXiv preprint arXiv:2305.15005.

Zhang, Y., Wang, M., Ren, C., Li, Q., Tiwari, P., Wang, B., et al. (2024). Pushing the limit of LLM capacity for text classification. arXiv preprint arXiv:2402.07470.

Zhong, P., Wang, D., & Miao, C. (2022). EEG-based emotion recognition using regularized graph neural networks. *IEEE Transactions on Affective Computing*, *13*(3), 1290–1301.