The DEEP Sensorium: a multidimensional approach to sensory domain labelling

Simona Corciulo[*]
Dipartimento di Informatica
University of Turin
Turin, Italy
simona.corciulo@unito.it

Livio Bioglio
Dipartimento di Studi Umanistici
University of Turin
Turin, Italy
livio.bioglio@unito.it

Valerio Basile
Dipartimento di Informatica
University of Turin
Turin, Italy
valerio.basile@unito.it

Viviana Patti
Dipartimento di Informatica
University of Turin
Turin, Italy
viviana.patti@unito.it

Rossana Damiano
Dipartimento di Informatica
University of Turin
Turin, Italy
rossana.damiano@unito.it

ABSTRACT

In this paper, we describe our intuitions about how language technologies can contribute to create new ways to enhance the accessibility of exhibits in cultural contexts by exploiting the knowledge about the history of our senses and the link between perception and language.

We evaluate the performance of five multi-class classification models for the task of sensory recognition and introduce the DEEP Sensorium (Deep Engaging Experiences and Practices - Sensorium), a multidimensional dataset that combines cognitive and affective features to inform systematic methodologies for augmenting exhibits with multi-sensory stimuli.

For each model, using different feature sets, we show that the features expressing the affective dimension of words combined with sub-lexical features perform better than uni-dimensional training sets.

CCS CONCEPTS

• **Human-centered computing → Accessibility technologies**; **Accessibility technologies**; • **Computing methodologies → Natural Language Processing**.

KEYWORDS

multidimensional lexicon, accessibility, affect, machine learning, multi-sensory design, museums

ACM Reference Format:
Simona Corciulo, Livio Bioglio, Valerio Basile, Viviana Patti, and Rossana Damiano. 2023. The DEEP Sensorium: a multidimensional approach to sensory domain labelling . In *Companion Proceedings of the ACM Web Conference 2023 (WWW '23 Companion), April 30–May 04, 2023, Austin, TX, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3543873.3587631

[*]All authors contributed equally to this research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
WWW '23 Companion, April 30–May 04, 2023, Austin, TX, USA
© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-9419-2/23/04…$15.00
https://doi.org/10.1145/3543873.3587631

1 INTRODUCTION

In the treatise *De Anima* (On the Soul) written in the c. 350 BC., Aristotle drew a hierarchy of the senses introducing vision as dominant compared to hearing, smell, touch and taste according to the strength of each sense in enabling people to experience the world and survive [28].

According to the Greek philosopher, humans are masters at perceiving through vision and hearing since culture and biology have affected how people act, communicate and think.

Today, the old distinction between vision and hearing and the remaining senses still supports a duality between two main groups: the powerful and noble senses [20, 21] – vision and hearing – described as the *distal* senses as opposed to the *proximal* ones (touch, taste, smell) [22].

The position of smell in the hierarchy has changed in the last decades: since many studies have reveled that vision has evolved to the detriment of the olfactory system, it has been demoted to a lower position after touch and taste [14, 28, 50]

Visual media dominate Western society since images are easy to access, create and understand [42]. The images surrounding us have a great visual appeal, and, as humans, we communicate our visually-dominant experiences with means that mirror the supremacy of gaze [29]. We use language and visual artefacts to materialise our intricate abstract thoughts in undoubtedly one of the most fascinating means people have used since the earlier graffiti painted at Altamira around 40,000 years ago [33].

Notwithstanding the supremacy of gaze and visual media, more than 1 billion people live with a type of disability for which a predominant part is affected by sensory diseases, and (17% affected by blindness or visual impairment, 6% affected by deafness or hearing loss) [23].

In museums, distal senses are the ones that are best emphasised since visitors are usually forbidden to touch artworks, putting a large portion of people with different sensory abilities at risk of being marginalised.

Evidence from the beginning of the XX century shows that granting multi-sensory access has been a challenge for curators for over a century. In 1913, the curator of the Museum of Sunderland, England, transformed its collections into a museum of touch [15]. Many photographs of the time show blind children with the taxidermy Wallance the lion, some specimens of crocodiles, fishes and other wonders.

Nowadays, thanks to the deeper knowledge of the brain and the sensory system of humans, the spread of innovative technologies provides ways to promote the use of senses in museums, preserving at the same time the artworks' integrity and promoting art enjoyment by giving people a chance to experience art like never before. Different approaches to encourage visitors to use their senses have been proposed during the last decades: many museums now provide some kind of tactile support to the blind by leveraging tactile replicas [9] or even 3D models of the artworks (e.g., statues) that can be touched with VR gloves [34].

Notwithstanding these progresses, designing a permanent multisensory exhibition for all is still very challenging.

The sensory augmentation of artworks, in fact, implies the extraction of sensory features from artworks and their mapping to technological devices [43].

However, the specificity of the stimulus-artwork association hinders the reuse of mapping knowledge across different artworks and contexts.

To overcome this limitation, we sustain that collecting knowledge from corpora of art-related texts could create a vast knowledge base less biased by subjective associations or specific contexts. The large availability of textual materials related to artworks in museums provides a way to explore sensory-based information stored in language in the art-related domain. Catalogue records, for example, are rich in sensory based information conveyed by single words, such as "soft" (tactile) or "vividness" (visual), and sometimes combined in highly evocative associations, such as "noisy stars" or "fresh tone" [10].

In this paper, we describe our intuitions about how language technologies can contribute to create new ways to enhance accessibility in museums by exploiting the link between perception and language.

In particular, we introduce a multi-dimensional lexical resource, the DEEP Sensorium (Deep Engaging Experiences and Practices - Sensorium), based on the alignment of existing resources that account for specific linguistic dimensions, such as perception and affect, and we test its use in the context of supervised learning. In particular, we address the task of assigning a sensory domain to single words, which lies at the basis of more complex linguistic tasks, such as the detection and interpretation of synaesthetic metaphors, and is crucial itself for grounding sensory design into linguistic evidence.

To do so, we evaluate the performance of five distinct multiclass classification models on verbal textual input using the DEEP Sensorium dataset to train them.

This paper is structured as follows. After surveying the related work on models of sensory information in language in Section 2, in Section 3 we introduce the DEEP Sensorium by describing its creation from existing sensory datasets and linguistic resources. Section 2 describes the experiments in sensory domain labeling conducted by training machine learning models on DEEP Sensorium. Results and limitations of the experiments are discussed in Section 4.3. Future work and Conclusion end the paper.

2 RELATED WORK

Perceptual information is crucial to understand how people process concrete and abstract concepts. In general, unlike the referents of an abstract concept (e.g. truth), concrete referents can be experienced through the senses (e.g. dog) [1, 6]. The two types of concepts differ since the concrete concepts are easier to learn, remember and process than the abstract ones [38, 46, 47]. To investigate how the brain processes conceptual knowledge, Binder et al. (2005) measured brain activation using the magnetic resonance imaging (MRI) on participants engaged in identifying concrete and abstract words and demonstrated that these cause different forms of activation in areas of the brain [3]. Compared to the concrete words, abstract words activate the areas connected to phonological and verbal working memory processes. An almost exclusive activation of the left hemisphere is observed for abstract words and areas related to executive functions of behaviour, such as problem-solving planning for concrete concepts [3]. Moreover, abstract concepts do not possess a single object as referent compared to concrete concepts [55]. They are more distant from perceptual modalities, varying more between contexts and demonstrating high values of semantic diversity [19]. In contrast, concrete concepts appear in a restricted and interrelated set of contexts with low values of semantic diversity [19]. And precisely because they are more heterogeneous than concrete concepts, abstract concepts are considered more complex, require more time for processing and can activate the emotional dimension [41].

Indeed, emotion and perception are two strongly connected dimensions. Emotions define how we perceive the world, organise our memory and make essential decisions facilitating both speed and probability for an information to be processed [44]. In some studies, the effects of abstractness may be determined by the higher emotional valence of abstract words [2]. [54] describe a preponderance of emotional features in the representation of abstract words. The effect could also be explained by considering the high level of semantic richness [54]. Considering a theoretical approach inspired by grounded cognition theory, representation and internal processing rely on the exact neural mechanisms as action and perception [40]. Abstract concepts cannot be embodied through sensory or motor information, and for this reason, internal affective experience can be sought as an alternative [40]. Zadra et al. (2011) argue that emotions exert a strong motivational influence on the environment because they provide immediate embodied information about the benefits and costs of an anticipated action [60].

Over the past decade, while interest in emotions has grown significantly, the relationships between sensory properties and emotions have served several purposes. Yuan and Barlow (2021) have studied how individuals respond to sensory information and how this is critical in facilitating online offers for humanised products [59]. Yand and Lee (2019) investigated the emotions aroused by food and drink or more straightforward taste and olfactory stimuli [58]. Globalisation is leading consumers to be exposed to new foods from very different countries around the world, pressing food industry to study a huge variety of products and identify factors influencing product acceptability from an affective perspective.

In Digital Humanities, the role of sensory information in non-literal, figurative language has recently attracted the attention of scholars. Su et al. [49] proposed a method for interpreting synaeshetic metaphors which relies on cross-modal similarity between different sensory modalities and affective features.

Similarly, Tekiroglu et al. [52] examined how sensory features affect the recognition of metaphors. They proposed a method to identify these correlations from a dependency-parsed corpus using a vocabulary that links English terms to sensory modalities. Their findings reveal that sensory features are essential for detecting metaphors.

As highlighted in [30], in spite of the fact that being able to automatically identify how the different sensory experiences are described would be very relevant to address different tasks, from emotion detection to metaphor identification, multi-sensory studies are still a niche area in natural language processing, with some very recent notable exceptions, such as the multilingual taxonomy about smell semi-automatically developed in [30], where olfactory terms in English, French, German and Italian have been extracted from different sources.

While the link between perception in language and emotions underpins many studies in the field of sensory marketing, rather than in cultural domain, this research ultimately aims to investigate how the knowledge about the linguistic correlates of sensory perception we have today can be applicable to langauge-based multi-sensory design in museums.

Given the support provided by the literature in favour of the combination of linguistic, perceptual and emotional information to learn and understand the meaning of words, our working assumption is that models that combine these different levels of cognition better reflect the acquisition of concepts and that this approach can be extended to the sensory meaning of words [5].

3 CREATING A MULTIDIMENSIONAL LINGUISTIC RESOURSE

Sensory lexicons for the English language are rare, often of modest size and created with highly variable sampling techniques. To overcome this limitation, in this paper we propose the DEEP Sensorium, a multidimensional dataset which combines cognitive and affective features. We decided to exploit different dimensions based on the evidence described in the literature surveyed above.

We created the DEEP Sensorium by mapping the knowledge stored in different lexicons. The features we employed relate to the knowledge of words and sensory domains, the link to emotions, and a data base on the values of age of acquisition, frequency, familiarity, imaginability, concreteness, and other variables.

3.1 Sensory dataset

Sensory resources rely on manual or automated annotation (the latter usually based on semantic similarity measures), with limitations documented in both cases [51, 56]. Nevertheless, expert human annotators could be the best alternative to achieve the data accuracy and quality required for training datasets. Manual labelling can successfully identify borderline cases that automated techniques cannot deal with. For these reasons, we decided to merge manually annotated datasets to create a larger collection of labeled data with the goal of attaining a sufficient data quality.

Lynott and Connell presented a dataset of 400 nouns and 423 adjective randomly sampled and annotated with perceptual strength ratings for the five traditional sensory modalities [24, 26]. Their study served to document the tendency of nouns to be associated with more modalities than adjectives. They showed that not all perceptual modalities are equally distinct and that correlations of varying degrees (non-existent, weak, moderate or strong) exist for pairs of modalities with a very strong positive relationship for the pairs smell and taste, sight and touch [26]. Both datasets also contains estimates of modality exclusivity, a value to describe whether or not a word belongs exclusively to a certain sensory modality.

Winter (2016) provided a dataset of 300 verbs consisting of perception verbs from the literature and random samples of words with a frequency above the median in the English Lexicon Project [57].

3.2 Affective lexicons

The use of emotional resources relies on to the widely acknowledged but intricate connection between senses and emotions. Senses have the capacity to elicit emotions, with a direct impact on the motivation and the behavior of human beings [17, 18, 27]. Conversely, in the figurative language, sensory-related words are widely employed to convey concepts or emotions that may be difficult to express literally [36, 61].

Emotional stimuli can be characterised by three dimensions: valence, dominance and arousal. Valence describes whether a stimulus is pleasurable (from pleasant to unpleasant), arousal describes the level of activation (from calm to aroused), dominance indicates the control exerted by the stimulus and can be used to discriminate emotions [53]. Among the three dimensions, dominance has been the least examined in the emotion literature, but it is estimated to be the one that varies the most between people [4, 37]. The three affective dimensions have been used to understand how people process emotional images or words, showing, for example, that valence has stronger effects for emotional images than for words [2].

Among the most widespread resources in sentiment analysis and emotion extraction, the National Research Council Valence Arousal Dominance Lexicon (NRC-vad) [31] and the National Research Council Emotion Intensity Lexicon (NRC-eil) [32] provide complementary information about the affective and emotional information conveyed by words. NRC-vad includes 20,000 English words and their scores for the three dimensions of meaning: valence, arousal and dominance. The values for the three dimensions were assigned manually using the best-worst scale. NRC-eil includes 5,814 English words with corresponding intensity scores for the eight emotions provided by the 'Plutchik Wheel' (anger, anticipation, disgust, fear, joy, sadness, surprise and trust) [12]. The intensity value for each word was annotated manually using the best-worst scale. For both, the scores range from 0 to 1.

3.3 MRC psycholinguistic database

The Medical Research Council Psycholinguistic Database (MRC Psycholinguistic Database) is one of the longest-running and most extensive dictionaries for selecting material for psycholinguistic tests and research in the field of artificial intelligence. In particular, it is recognised as an essential source from which to draw data that can be used in the design of models for natural language processing [8]. The psycholinguistic database contains many concepts closely related to the sensory domain, such as imaginability, concreteness and abstractness: it has been shown that such words are related to perception, are very concrete and, therefore, easy to imagine [46].

The MRC Database consists of 150, 837 words annotated with 26 linguistic attributes; it differs from other dictionaries for the absence of semantic information. We decided to include several of these attributes to our dataset, but future studies may investigate more in detail the relevance of such attributes over the other ones, in order to purge the attributes that show low impact for the field of artificial intelligence.

3.4 The DEEP Sensorium

The DEEP Sensorium was created by aligning sensory data about single words from the lexicons described above, with the goal of creating a multidimensional resource, in line with the evidence from cognitive studies on word learning and understanding surveyed in Section 2. Starting from the existing collections of sensory data, namely those provided by [25, 26] and [57], we first merged these resources by obtaining a sensory dataset of 823 items, then we aligned this dataset with the psycholinguistic information contained in MRC psycholinguistic database and the affective information contained in NRC-vad [53] and NRC-eil [32].

Emotion	No.	Domain	No.	W. Type	No.
Joy	147	Vision	551	Noun	305
Trust	138	Auditory	118	Verb	242
Sadness	122	Haptic	57	Adjective	160
Fear	89	Taste	47	Adverb	41
Anger	85	Smell	13	0ther	26
Disgust	84			Past participle	10
Anticipation	74			Interjection	1
Surprise	47			Rare	1

Table 1: Number of entries for each emotion, sensory domain and type of word in the DEEP Sensorium dataset

The DEEP Sensorium contains 25 features, inherited from the resources from which it has been created; it is stored as a comma-separated values (CSV) file[1].

A group of features is mostly related to measures of word frequency (e.g. Krucera Francis written frequency [48], Thorndike-Lorge written frequency [16], Brown verbal frequency [7]) and grammar (e.g number of syllables, common part of speech, pronunciation variability, irregular plural) extracted from the MRC psycholinguistic database; a second group of features encompasses the experimentally collected response to the single words (e.g., concreteness, abstractness, familiarity [8], imaginability [8], meaningfulness [8, 39]) and their affective value (valence, arousal, dominance, intensity, emotion) extracted from both MRC and NRC (vad and eil).

[1]Available on request

Domain	joy	tru	fea	sur	sad	dis	ang	ant	TOT
Vision	**52**	50	31	18	36	22	23	33	265
Auditory	**12**	7	11	2	7	6	**12**	5	62
Haptic	3	3	3	-	**7**	5	3	-	24
Taste	**5**	3	1	2	2	4	1	2	20
Smell	-	1	1	-	-	**5**	-	-	7
TOTAL	**72**	64	47	22	52	42	39	40	378

Table 2: Number of emotions for each domain in the DEEP Sensorium dataset, where we consider only the 229 unique entries: the total number of emotions is bigger because an entry may convey several emotions

Finally, a third group of features is related to the sensory domain of words.

We decided to exclude from the DEEP Sensorium some features contained in the MRC psycholinguistic database (such as phonetic transcription and stress pattern) since only a very limited subset of entries reported a value for these features.

The lexicons used differ significantly in size:

- NRC-eil contains 9829 entries,
- NRC-vad contains 19971 entries,
- MRC psycholinguistic contains 150,837 instances.

Clearly, the different design of these resources and their differences in size pose problems for the alignment.

Concerning the overlap with the affective lexicons, the merged sensory dataset shared 378 entries with the NRC-eil and 709 with the NRC-vad.

As as result, many of the words in the merged sensory dataset had to be excluded (455 entries) from the DEEP Sensorium dataset in order to create a training set as complete as possible for all features.

However, the final size of the dataset is larger because in the MRC psycholinguistic database some entries appear more than once, since the same word can appear several times with different configurations of values (e.g., different acquisition ages, different abstractedness values, and even different parts of speech), differently from standard lexicons, where entries are not repeated (unless in case of ambiguity, as in the case of 'bad' noun and 'bad' adjective). We decided to keep the status of each word separate for all parts of speech to have more control over the annotation regarding psycholinguistic variables. So, at the end of the alignment process, the DEEP Sensorium contained 786 entries and 229 unique entries.

It is worth stressing that the distribution of values over the single features is the DEEP Sensorium is not balanced, in particular for what concerns sensory and affective features, as represented in Table 1. This is not entirely unexpected for sensory domains, since the unbalanced distribution of domains mirrors the influence of perception on language and vice versa. Unlike the proximal senses, in fact, the distal senses are the senses humans use the most. Consequently, in communicating our experiences of the world through language, the vocabulary associated with sight and hearing is certainly dominant. The distribution shows that the proximal senses are poorly represented, with insufficient instances for the olfactory domain (Table 1, Domain). Similarly, the distribution of emotions is not balanced: we can highlight a very high representation of Joy and Trust (Table 1, Emotion).

The hierarchy of senses describes the evolutionary advantage of some senses in gathering much information from the environment. The distal senses benefit humans: by vision and hearing, distant stimuli in space are gathered, activating emotions beneficial to avoid health risks. Several empirical studies have detected that many emotions positively correlated with risk aversion, including fear, happiness, anger and surprise [35]. Examining the data (Table ??), we identified some emotions elicited by risk aversion for the distal senses, except for surprise related to the hearing domain. In this view, we are led to speculate that some words might be underused in a language not only because they relate to experiences challenging to capture or process by people (e.g. smells) but because they need to be functional for processing a specific emotion. Thus, insufficient samples could mean no links between sensory words and the possibility of activating an emotional response such as that provoked by risk aversion. The distribution of emotions in the sensory domains can be partly explained by recourse to the literature. Concerning the haptic-sadness-disgust correlation, it has been shown that haptic effectively communicates the arousal and emotions of happiness, sadness, anger and fear [13]. In contrast, less attention has been given to communicating disgust and surprise. More recently, through the study of haptic communication of emotions, it has been found that people can identify anger, disgust, fear, gratitude, happiness, love, sadness and sympathy from the experience of being touched on the arm or body by a stranger, without seeing the touch [17, 18]. These results could support the relatively more even distribution of samples among the vision and hearing domains: the two senses are those that manly alert people for reasons related to the risk aversion.

Regarding the link between taste and joy, for theoretical and empirical reasons, it seems likely that different emotions influence eating in specific ways. Basic emotions such as anger, fear and sadness have distinct motivational functions: when we experience joy, the motivation to eat to enjoy food increases, whereas negative emotions increase the tendency to eat to cope with the negative emotional state [27]. Regarding the connection between odour and fear, many studies suggest that humans can become fearful after exposure to olfactory fear signals [11]. Independent of visual and auditory information, olfactory fear signals produced by senders induce fear in receivers outside of conscious access. These results contrast the traditional view that emotions are communicated exclusively through visual and linguistic channels.

In order to study the contribution of the single sources of information (psycholinguistic and affective) to the sensory labelling task, in the experimental setting we availed ourselves of different aggregations of features obtained by aligning the merged sensory dataset with the single lexical resources.

The NRC-eil subset has been obtained by aligning the sensory lexicon with the NRC-eil; it contains 229 entries annotated for emotion and intensity.

The NRC-vad subset has been obtained by aligning the sensory lexicon with the NRC-vad; it contains 708 entries annotated for valence, dominance, arousal, and dominance.

The psycholinguistic subset has been obtained by aligning the sensory lexicon with the features extracted from the MRC psycholinguistic database. This subset contains 1828 instances annotated for 27 features including frequency, grammar and processing information, as mentioned above.

4 EXPERIMENTAL EVALUATION

In the experimental phase, we tested multi-class classification models to study how they perform in the sensory labeling task, namely the task of predicting the sensory domain for a single verbal input from the other features. The models were trained using the DEEP Sensorium multidimensional dataset described in the previous section. In order to evaluate the capabilities of our resource compared to the ones employed for its creation (see Section 3), different feature sets (each corresponding to one of the original resources) were used to evaluate how they impact the performance of the models.

4.1 Models

We employed five different multi-class classification models to explore the automation of the sensory modeling task. Many of these models generally support binary classification, but specific extensions allow them to be used as multi-binary classifiers for multi-class classification problems [45].

The choice of opting for multi-class classification over multi-label classification depends on the expected applications of the models. In our case, opting for classification in which the input is associated with a single class (a single sensory domain), could serve tasks such as synesthetic metaphor detection [10, 49] and interpretation, in which assigning multiple classes to each input might be not crucial to the success of the task. In contrast, multi-label classification could contribute to the investigation of perceptual experience through written and spoken language in all its complexity.

The following models were employed:

- Logistic Regression (LR),
- Decision Trees (DT),
- Random Forest (RF),
- Support Vector Machine (SVM),
- K-Nearest Neighbors (KNN).

4.2 Experimental setting

The preprocessing stage works as follows. Each categorical feature is encoded using one-hot encoding: since the number of categories is quite low, such process does not dramatically increase the dimensional space. Words are converted into vectors of values using a word embedding approach: we employ pre-trained GloVe vectors, trained on the English-language Wikipedia 2014 + Gigaword 5 corpus [2]. In order to keep the dimensional space reasonable, we decided to use vectors with 50 dimensions. The number of features of each sample after the preprocessing stage is 103.

The hyperparameter optimization step is performed with a Grid Search approach, which simply executes a complete search over a given subset of the hyperparameters space of the training algorithm. The set of hyperparameters we decided to explore varies according to the ML algorithm: for Logistic Regression, we explored the norm of the penalty and the inverse of regularization strength; for Decision Tree and Random Forest, we inspected the minimum number of samples required to be at a leaf node and to split an

[2] Available at nlp.stanford.edu/projects/glove/.

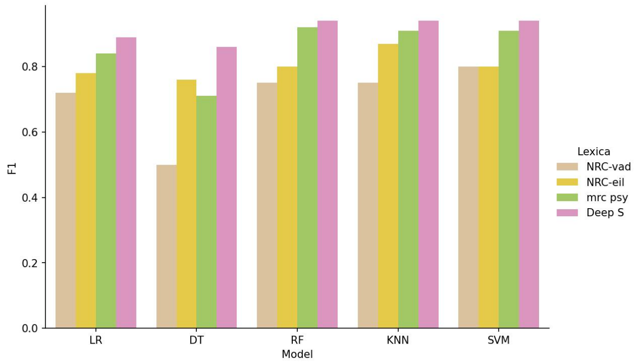

Figure 1: Value of F1 on each model trained on the DEEP Sensorium dataset and its subsets described in Section 3.4

internal node, the maximum depth of the tree and the function to measure the quality of a split; for KNN, the number of neighbors, the weight function and the metric used for distance computation; for SVM, the regularization parameter and the kernel type.

A 10-fold cross-validation was employed with a 80% for training and 20% for testing.

4.3 Results and discussion

Figure 1 depicts the value of F1 score calculated for each model and trained with each resource, namely, the multi-dimensional DEEP Sensorium and the uni-dimensional subsets (NRC-eil subset, NRC-vad subset, psycholinguistic subset) described in Section 3. The figure shows that the models trained with our multi-dimensional resource always outperform the models trained with the uni-dimensional subsets of features, with an F1 score close to or greater than 0.9.

The superiority of the multi-dimensional model is in line with the expectations set by the psycholinguistic literature that has inspired the creation of the DEEP Sensorium dataset (see Section 3).

As described in the literature, sensory information is crucial for understanding concrete and abstract concepts, so the knowledge about specific features, such as abstractness and concreteness, can facilitate sensory classification. So, even if the psycholinguistic features alone (the psycholinguistic subset) have a slightly worse performance than the multi-dimensional dataset (F1 weighted score ≥ 0.73 with values between 0.73 and 0.92 depending on the models versus F1 weighted scores ≥ 0.72 with a range of values between 0.87 and 0.94), their contribution is probably more relevant than that of the affective subsets.

Although the link between emotions and perception is fundamental, as shown by research in perception and affect, the results obtained by the affective subsets suggest that the contribution of dimensional models, represented here by the NRC-vad subset, is smaller than the one provided by categorical models, namely the NRC-eil subset. Both subsets, in fact, perform worse than the DEEP Sensorium dataset, with a larger gap for the NRC-vad subset (F1 weighted score ≥ 0.54 with values between 0.54 and 0.80) than the NRC-eil subset (≥ 0.76, with values ranging from 0.76 to 0.86). This confirms the findings reported in the literature, where the role of valence, insufficiently studied, seems more relevant for images than words.

Model	Precision	Recall	F1-score
LT	0.90	0.89	0.89
DT	0.89	0.86	0.87
RF	**0.95**	**0.94**	**0.94**
KNN	0.94	**0.94**	**0.94**
SVM	0.94	**0.94**	**0.94**

Table 3: Precision, Recall and F1 score for each model trained on DEEP Sensorium dataset.

Once stated that our dataset guarantees in general better performances in the sensory domain labeling task with respect to the other resources, we evaluate the performances of each model trained on DEEP Sensorium. Table 3 reports the value of Precision, Recall and F1 (calculated with the weighted approach) of each model: such results show that Random Forest, Support Vector Machine and K-Nearest Neighbors are the best models, with more or less the same performances. The F1 weighted score is 0.94 for all these three models; recall is 0.94 for all, and precision ranges from 0.94 (Support Vector Machine and K-Nearest Neighbors) to 0.95 (Random Forest).

For understanding more in detail the performance of the three best models, we analyse the metrics on each class: Table 4 contains the values of Precision, Recall and F1 for each sensory domain calculated on the best models trained with DEEP Sensorium dataset. We notice that in general the Vision domain (the distal domain par excellence) exhibits the best scores across all the models (with KNN, precision 0, 97, recall 1.0, F1 0.98), probably due to the larger number of samples in our dataset, while all the models perform a bit poorly on Haptic domain (one of the proximal domains), with the worst performance for the SVM model (precision 0.86, recall 0.55, F1 0.67). Even if the relevance of the Haptic domain may seem limited, this domain is of great importance for multi-sensory design, as witnessed by the experiments where visual features, inaccessible to some users, have been mapped onto relatively inexpensive haptic devices, as described in Section 1.

Finally, Random Forest is the model with the most balanced performances on all the sensory domains, while on the opposite side Support Vector Machine shows the most unbalanced scores, with F1 ranging from 0.67 for the Haptic domain to a perfect score for the Smell domain.

4.4 Limitations

Sensory information in language mirrors the position of the senses in Aristotle's hierarchy and is reflected in the severe imbalance of the linguistic resources along the sensory dimension. Our dataset is not exempt from this imbalance, privileging distal senses (Vision and Hearing) over proximal senses (Smell, Taste and Haptic). As a consequence, the already low availability of sensory data about words is even more dramatic for the lower domains in the hierarchy, as reported in Section 1.

In this work, we tested five different models for multi-class classification and used weighted classes to avoid oversampling, which dramatically increases the dimension of the dataset, and undersampling, which is unsuitable for small datasets.

		Precision			Recall			F1-score		
Domain	**No. samples**	**RF**	**KNN**	**SVM**	**RF**	**KNN**	**SVM**	**RF**	**KNN**	**SVM**
Vision	70.25%	0.93	0.97	0.93	1.00	1.00	1.00	0.97	0.98	0.97
Auditory	15.19%	1.00	0.88	1.00	0.88	0.88	0.92	0.93	0.88	0.96
Haptic	6.96%	1.00	0.88	0.86	0.64	0.64	0.55	0.78	0.74	0.67
Taste	5.70%	1.00	1.00	1.00	0.78	0.78	0.78	0.88	0.88	0.88
Smell	1.90%	0.75	0.75	1.00	1.00	1.00	1.00	0.86	0.86	1.00
weighted average		0.95	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94

Table 4: Precision, Recall and F1 score for each sensory domain in the models Random Forest (RF), K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) trained on DEEP Sensorium dataset

The imbalanced distribution of the data over the sensory domains is even exacerbated when it comes to assessing the role of the single dimensions (affective and psycholinguistic) for sensory labeling task. The subsets employed to measure the impact of the single dimensions (affective and psycholinguistic) have different sizes and features. For these reasons, we have concluded that we do not control the risk of overfitting and that the issue of how to deal with the intrinsic imbalance of linguistic data is still open.

Last, but not least, the merged sensory datasets which constitute the core of the DEEP Sensorium have been created with specific applications in mind in the field of product design and food industry, so they may reflect the characteristics of these domains. The artistic domain, which is the ultimate target of our work, may reflect a different role and distribution of sensory information, a risk that can be assessed only by extending the approach to data extracted from art-related texts such as the initial dataset presented by [10].

5 CONCLUSION AND FUTURE WORKS

In this paper, we introduced a new resource for the assignment of sensory domain in language which combines sensory data with affective and pyscholinguistic information from existing resources into a novel merged resource, the DEEP Sensorium dataset.

To evaluate the assumption, suggested by the literature, that learning the meaning of words requires the integration of linguistic, perceptual and emotional information, we tested our resource with a set of multi-dimensional classification models. The results confirm that a multidimensional approach yields better results than the single dimensions taken separately from each other.

Among the classification model we employed in our research, Random Forest (RF), K-Nearest Nighbors (KNN) and Support Vector Machine (SVM) showed the best performances, with the same value for the F1 score: analyzing the detailed metrics of each single class (the sensory domains) we noticed that Random Forest exhibits the most balanced performances on all the classes, while the other two models are slightly biased on some sensory domain.

For future work, our aim is to measure the effect of narrower feature subsets by scaling the dataset and integrating data from objective measurements of engagement, such as brain activation, and data from the artistic domain. These two steps, in fact, are crucial to achieve the objective of supporting multi-sensory design in the cultural sector. In addition, the current coverage of the resource should be extended, and part of future research will focus on integrating new annotated words.

Furthermore, in the future, we plan to test the performance for more complex classification tasks, such as multi-label classification.

REFERENCES

[1] Lawrence W. Barsalou and Katja Wiemer-Hastings. 2005. Situating Abstract Concepts. In *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking*, Diane Pecher and Rolf A.Editors Zwaan (Eds.). Cambridge University Press, Cambridge, 129–163. https://doi.org/10.1017/CBO9780511499968.007
[2] Mareike Bayer and Annekathrin Schacht. 2014. Event-related brain responses to emotional words, pictures, and faces–a cross-domain comparison. *Frontiers in psychology* 5 (2014), 1106.
[3] Jeffrey R Binder, Chris F Westbury, Kristen A McKiernan, Edward T Possing, and David A Medler. 2005. Distinct brain systems for processing concrete and abstract concepts. *Journal of cognitive neuroscience* 17, 6 (2005), 905–917.
[4] Margaret M Bradley and Peter J Lang. 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry* 25, 1 (1994), 49–59.
[5] Elia Bruni, Nam-Khanh Tran, and Marco Baroni. 2014. Multimodal distributional semantics. *Journal of artificial intelligence research* 49 (2014), 1–47.
[6] Marc Brysbaert, Amy Beth Warriner, and Victor Kuperman. 2014. Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior research methods* 46 (2014), 904–911.
[7] Curt Burgess and Kay Livesay. 1998. The effect of corpus size in predicting reaction time in a basic word recognition task: Moving on from Kučera and Francis. *Behavior Research Methods, Instruments, & Computers* 30, 2 (1998), 272–277.
[8] Max Coltheart. 1981. The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A* 33, 4 (1981), 497–505.
[9] Radu Comes. 2016. Haptic devices and tactile experiences in museum exhibitions. *Journal of Ancient History and Archaeology* 3, 4 (2016), 60–64.
[10] Simona Corciulo and Viviana Patti ad Damiano Rossana. 2022. Towards the Construction of a Dataset of Art-Related Synaesthetic Metaphors: Methods and Results. In *Proceedings of the 1st Workshop on Artificial Intelligence for Cultural Heritage, AI4CH 2022, co-located with the 21st International Conference of the Italian Association for Artificial Intelligence (AIxIA 2022), Udine, Italy, November 28, 2022.* CEUR Workshop Proceedings, Udine, Italy, 113–125. http://ceur-ws.org/Vol-3286/12_paper.pdf
[11] Jasper HB De Groot, Gün R Semin, and Monique AM Smeets. 2014. I can see, hear, and smell your fear: comparing olfactory and audiovisual media in fear communication. *Journal of Experimental Psychology: General* 143, 2 (2014), 825.
[12] Melissa Donaldson. 2017. Plutchik's wheel of emotions—2017. Update.
[13] Mohamad A Eid and Hussein Al Osman. 2015. Affective haptics: Current research and future directions. *IEEE Access* 4 (2015), 26–40.
[14] Yoav Gilad, Victor Wiebe, Molly Przeworski, Doron Lancet, and Svante Pääbo. 2004. Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. *PLoS biology* 2, 1 (2004), e5.
[15] Ian Grosvenor and Natasha Macnab. 2013. Seeing through touch: the material world of visually impaired children. *Educar em Revista* 49 (2013), 39–57.
[16] John F Hall. 1954. Learning as a function of word-frequency. *The American journal of psychology* 67, 1 (1954), 138–140.
[17] Matthew J Hertenstein, Rachel Holmes, Margaret McCullough, and Dacher Keltner. 2009. The communication of emotion via touch. *Emotion* 9, 4 (2009), 566.
[18] Matthew J Hertenstein, Dacher Keltner, Betsy App, Brittany A Bulleit, and Ariane R Jaskolka. 2006. Touch communicates distinct emotions. *Emotion* 6, 3 (2006), 528.
[19] Paul Hoffman, Matthew A Lambon Ralph, and Timothy T Rogers. 2013. Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior research methods* 45 (2013), 718–730.

[20] Martin Jay. 1993. 1. The Noblest of the Senses: Vision from Plato to Descartes. In *Downcast Eyes*. University of California Press, Berkeley, 21–82. https://doi.org/10.1525/9780520915381-003

[21] Danijela Kambaskovic-Sawers and Charles T Wolfe. 2014. The senses in philosophy and science: from the nobility of sight to the materialism of touch.

[22] Carolyn Korsmeyer. 2019. A Tour of the Senses. *The British Journal of Aesthetics* 59, 4 (2019), 357–371.

[23] Gloria L Krahn. 2011. WHO World Report on Disability: a review. *Disability and health journal* 4, 3 (2011), 141–142.

[24] George Lakoff. 2008. The neural theory of metaphor. In *The Cambridge Handbook of Metaphor and Thought*, Raymond W.Editor Gibbs, Jr. (Ed.). Cambridge University Press, Cambridge, 17–38. https://doi.org/10.1017/CBO9780511816802.003

[25] Dermot Lynott and Louise Connell. 2009. Modality exclusivity norms for 423 object properties. *Behavior Research Methods* 41, 2 (2009), 558–564.

[26] Dermot Lynott and Louise Connell. 2013. Modality exclusivity norms for 400 nouns: The relationship between perceptual experience and surface word form. *Behavior research methods* 45, 2 (2013), 516–526.

[27] Michael Macht, S Roth, and Heiner Ellgring. 2002. Chocolate eating in healthy men during experimentally induced sadness and joy. *Appetite* 39, 2 (2002), 147–158.

[28] Asifa Majid, Seán G Roberts, Ludy Cilissen, Karen Emmorey, Brenda Nicodemus, Lucinda O'grady, Bencie Woll, Barbara LeLan, Hilário De Sousa, Brian L Cansler, et al. 2018. Differential coding of perception in the world's languages. *Proceedings of the National Academy of Sciences* 115, 45 (2018), 11369–11376.

[29] M. McLuhan. 1962. *The Gutenberg Galaxy: The Making of Typographic Man*. University of Toronto Press, Toronto. https://books.google.it/books?id=y4C644zHCWgC

[30] Stefano Menini, Teresa Paccosi, Serra Sinem Tekiroğlu, and Sara Tonelli. 2022. Building a Multilingual Taxonomy of Olfactory Terms with Timestamps. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France, 4030–4039. https://aclanthology.org/2022.lrec-1.429

[31] Saif Mohammad. 2018. Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Melbourne, Australia, 174–184. https://doi.org/10.18653/v1/P18-1017

[32] Saif Mohammad. 2018. Word Affect Intensities. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA), Miyazaki, Japan, 174–183. https://aclanthology.org/L18-1027

[33] Pilar Fatás Monforte. 2021. The Cave of Altamira and Modern Artistic Creation. In *Visual Culture, Heritage and Identity: Using Rock Art to Reconnect Past and Present*. Archaeopress Publishing Ltd, Oxford, UK, 24.

[34] Konstantinos Moustakas, Michael G Strintzis, Dimitrios Tzovaras, Sébastien Carbini, Olivier Bernier, Jean-Emmanuel Viallet, Stephan Raidt, Matei Mancas, Mariella Dimiccoli, Enver Yagci, et al. 2006. Masterpiece: physical interaction and 3D content-based search in VR applications. *IEEE MultiMedia* 13, 3 (2006), 92–100.

[35] Yen Nguyen and Charles N Noussair. 2014. Risk aversion and emotions. *Pacific economic review* 19, 3 (2014), 296–312.

[36] Andrew Ortony and Lynn Fainsilber. 1987. The role of metaphors in descriptions of emotions. In *Theoretical Issues in Natural Language Processing 3*.

[37] Charles Egerton Osgood, George J Suci, and Percy H Tannenbaum. 1957. *The measurement of meaning*. University of Illinois press, Champaign, Illinois, US.

[38] Allan Paivio. 1991. Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue canadienne de psychologie* 45, 3 (1991), 255.

[39] Allan Paivio, John C Yuille, and Stephen A Madigan. 1968. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of experimental psychology* 76, 1p2 (1968), 1.

[40] Marina Palazova, Werner Sommer, and Annekathrin Schacht. 2013. Interplay of emotional valence and concreteness in word processing: An event-related potential study with verbs. *Brain and language* 125, 3 (2013), 264–271.

[41] Sophie Pauligk, Sonja A Kotz, and Philipp Kanske. 2019. Differential impact of emotion on semantic processing of abstract and concrete words: ERP and fMRI evidence. *Scientific Reports* 9, 1 (2019), 14439.

[42] Sarah Pink. 2006. *The future of visual anthropology: Engaging the senses*. Routledge, London, UK.

[43] Tom Pursey and David Lomas. 2018. Tate Sensorium: An experiment in multisensory immersive design. *The Senses and Society* 13, 3 (2018), 354–366.

[44] Stanley Rachman. 1980. Emotional processing. *Behaviour research and therapy* 18, 1 (1980), 51–60.

[45] Mahendra Sahare and Hitesh Gupta. 2012. A review of multi-class classification for imbalanced data. *International Journal of Advanced Computer Research* 2, 3 (2012), 160.

[46] Paula J Schwanenflugel. 2013. Why are abstract concepts hard to understand? In *The psychology of word meanings*. Psychology Press, London, UK, 235–262.

[47] Paula J Schwanenflugel and Edward J Shoben. 1983. Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 9, 1 (1983), 82.

[48] Bernard J Shapiro. 1969. The subjective estimation of relative word frequency. *Journal of verbal learning and verbal behavior* 8, 2 (1969), 248–251.

[49] Chang Su, Xiaomei Wang, Zita Wang, and Yijiang Chen. 2019. A model of synesthetic metaphor interpretation based on cross-modality similarity. *Computer Speech & Language* 58 (2019), 1–16.

[50] Robert W Sussman, D Tab Rasmussen, and Peter H Raven. 2013. Rethinking primate origins again. *American Journal of Primatology* 75, 2 (2013), 95–106.

[51] Serra Sinem Tekiroğlu, Gözde Özbal, and Carlo Strapparava. 2014. Sensicon: An automatically constructed sensorial lexicon. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 1511–1521.

[52] Serra Sinem Tekiroğlu, Gözde Özbal, and Carlo Strapparava. 2015. Exploring sensorial features for metaphor identification. In *Proceedings of the Third Workshop on Metaphor in NLP*. Association for Computational Linguistics, Denver, Colorado, 31–39. https://aclanthology.org/W15-1404

[53] Gyanendra K Verma and Uma Shanker Tiwary. 2014. Multimodal fusion framework: A multiresolution approach for emotion classification and recognition from physiological signals. *NeuroImage* 102 (2014), 162–172.

[54] Gabriella Vigliocco, Lotte Meteyard, Mark Andrews, and Stavroula Kousta. 2009. Toward a theory of semantic representation. *Language and Cognition* 1, 2 (2009), 219–247.

[55] Caterina Villani, Luisa Lugli, Marco Tullio Liuzza, Roberto Nicoletti, and Anna M Borghi. 2021. Sensorimotor and interoceptive dimensions in concrete and abstract concepts. *Journal of memory and language* 116 (2021), 104173.

[56] Bodo Winter. 2016. *The sensory structure of the English lexicon*. University of California, Merced.

[57] Bodo Winter. 2016. *The sensory structure of the English lexicon*. University of California, Merced, Merced, US.

[58] Jiyun Yang and Jeehyun Lee. 2019. Application of sensory descriptive analysis and consumer studies to investigate traditional and authentic foods: A review. *Foods* 8, 2 (2019), 54.

[59] Lingyao Yuan and Jordan Barlow. 2021. Sensitive to the Digital Touch? Exploring Sensory Processing Sensitivity and Its Impact on Anthropomorphized Products in E-Commerce. In *54th Hawaii International Conference on System Sciences, HICSS 2021, Kauai, Hawaii, USA, January 5, 2021*. ScholarSpace, Kauai, Hawaii, US, 1–10.

[60] Jonathan R Zadra and Gerald L Clore. 2011. Emotion and perception: The role of affective information. *Wiley interdisciplinary reviews: cognitive science* 2, 6 (2011), 676–685.

[61] Jie Zhou, Qi Su, and Pengyuan Liu. 2020. A metaphorical analysis of five senses and emotions in mandarin Chinese. In *Chinese Lexical Semantics: 20th Workshop, CLSW 2019, Beijing, China, June 28–30, 2019, Revised Selected Papers 20*. Springer, 607–617.