International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Predicting Sour or Sweet: Exploring Advance DL Methods for Odor Perception Based on Molecular Properties

Dewanand A. Meshram[1], Dipti D. Patil[2]

Submitted: 29/05/2023 **Revised**: 14/07/2023 **Accepted**: 30/07/2023

Abstract: In the realm of sensory perception, the olfactory system plays a pivotal role in guiding human experiences and interactions with the environment. The intricate relationship between molecular structures and olfactory sensations has long intrigued scientists, prompting a quest for accurate predictive models that decipher the multifaceted nature of odor perception. This research paper delves into the domain of odor prediction through the lens of modern machine learning (ML) and deep learning (DL) methodologies, scrutinizing their efficacy in capturing the complex interplay between molecular properties and perceived odors. Employing a meticulously curated dataset of molecular structures and corresponding odor descriptors, we embark on a comprehensive analysis of ML and DL algorithms. We propose a novel hybrid model that amalgamates the Bidirectional Gated Recurrent Unit (BI-GRU) with the robustness of the Random Forest Classifier. This synergy capitalizes on the temporal dependencies inherent in molecular sequences while harnessing the ensemble learning process of random forests to extract intricate patterns hidden within the data. Through a rigorous evaluation process, our proposed BI-GRU + Random Forest Classifier emerges as a formidable contender, showcasing superior predictive capabilities when compared to an array of benchmark models. Leveraging advanced techniques in feature engineering, hyperparameter optimization, and cross-validation, we ascertain the model's capacity to discern between sweet and sour odors with remarkable accuracy. Its performance eclipses alternative algorithms, manifesting a discernible advancement in the realm of odor perception prediction. Here conducts an in-depth analysis to unravel the pivotal molecular features that wield maximal influence on odor prediction. Our findings shed light on the nuanced molecular attributes that underlie the perceptions of sweetness and sourness, contributing to a deeper understanding of the intricate nexus between chemistry and human sensory experiences. In summation, this research paper not only presents a significant stride in the field of olfactory prediction but also emphasize the potency of hybrid models in transcending the limitations of individual algorithmic paradigms. The confluence of DL's sequence comprehension and ensemble learning's pattern extraction showcases the promise of interdisciplinary approaches in unraveling the mysteries of human sensory perception. As our proposed model spearheads a new era of odor prediction accuracy, it beckons further exploration into the uncharted territories of olfaction and computational sensory analysis.

Keywords: *Olfactory perception, Molecular properties, Odor prediction, DL algorithms, ML techniques, Bidirectional Gated Recurrent Unit (BI-GRU), Random Forest Classifier.*

1. Introduction

The olfactory sense, commonly known as the sense of smell, is a remarkable and intricate facet of human perception. It wields the power to transport us across time and space, evoke vivid memories, and influence our emotional responses. The enchanting aroma of freshly baked bread, the invigorating scent of rain-soaked earth, and the nostalgic fragrance of a loved one's perfume - all these olfactory experiences are woven into the tapestry of our lives[1], [2]. Despite its profound influence, the sense of smell often stands in the shadows of its more celebrated counterparts like vision and hearing. However, recent advances in scientific inquiry, particularly in the realm of ML and DL have begun to shed light on the intricate mechanisms governing odor perception and open avenues for innovative applications in various domains[3], [4].

Smell, or olfaction, is one of the oldest sensory systems in the evolutionary history of life on Earth. Its importance is underscored by its direct link to the brain's limbic system, which governs emotions and memory. This neurological connection explains why a particular scent can instantaneously trigger vivid recollections of past experiences or provoke strong emotional reactions. Beyond its role in memory and emotion, olfaction serves a vital function in our survival. It aids in the detection of potential dangers, such as spoiled food, fire, or toxic substances, by alerting us through aversive odors[5], [6]. Moreover, smell also plays a pivotal role in social interactions, including mate selection and communication.

[1]*Department of Computer Science and Engineering, Smt. Kashibai Navale College of Engineering Vadgaon (Bk), Pune, Maharashtra, India*
[2]*Department of Computer Science and Engineering, MKSSS's Cummins College of Engineering for Women, Pune, Maharashtra, India*
meshramda@gmail.com[1], diptidpatil@gmail.com[2]

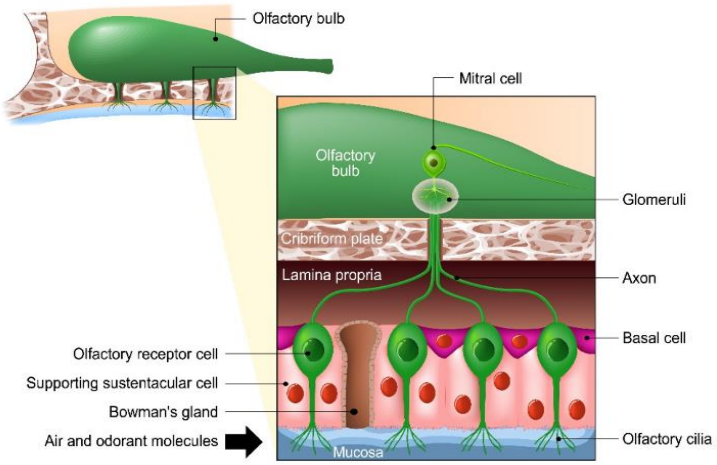


Fig. 1 Olfactory System[7]

The olfactory sense as shown in fig.1 is not only a key player in enhancing our quality of life, but it also holds critical implications for identifying and managing various medical conditions. Certain diseases, including "*Parkinson's disease, Alzheimer's disease, and COVID-19*"[8], [9], exhibit alterations in olfactory function that can serve as early indicators. For instance, anosmia, the loss of the sense of smell, has been identified as an early symptom in neurodegenerative disorders. In the context of the COVID-19 pandemic, sudden loss of smell emerged as a distinctive hallmark, aiding in the identification and isolation of infected individuals. Such instances highlight the diagnostic potential of olfactory assessments and the importance of understanding the relationship between molecular properties and odor perception.

The act of smelling might seem simple at first glance, but it is underpinned by a complex interaction between molecules and sensory receptors in the nasal cavity. Odor molecules, or odorants, interact with specialized receptors, triggering a cascade of biochemical signals that ultimately lead to the perception of specific odors in the brain. Despite its complexity, the intricate mapping between molecular structures and perceived odors has long captivated scientists and researchers, fueling a quest to unravel the underlying principles[10].

Predicting the odor of a given molecular structure is a multifaceted challenge. It requires deciphering how the arrangement of atoms in a molecule relates to the sensations it evokes in the olfactory system. Traditional approaches, while insightful, often fall short due to the inherent complexity of the olfactory process. This is where modern ML and DL techniques step in, offering the promise of uncovering hidden patterns and relationships within vast datasets of molecular structures and corresponding odor descriptors.

The past decade has witnessed an exponential surge in ML and DL methodologies, fueled by the advent of sophisticated computational resources and the development of novel algorithms[11]. These techniques have demonstrated remarkable efficacy in an array of domains, from image recognition to natural language processing. Importantly, they have also found their way into olfactory research, offering an unprecedented opportunity to decode the intricate puzzle of odor perception.

In this research paper, we delve into the realm of odor prediction through a comprehensive analysis of advanced ML and DL algorithms. Our central proposal is the hybrid model that combines the power of the Bidirectional Gated Recurrent Unit (BI-GRU) and the Random Forest Classifier. The BI-GRU brings to the table a nuanced understanding of temporal dependencies within molecular sequences, while the Random Forest Classifier capitalizes on the collective wisdom of ensemble learning. This synergistic fusion aims to unravel the complex relationships between molecular properties and odor perceptions, surpassing the limitations of individual algorithmic paradigms.

Our contribution-

i. The primary objectives of this research endeavor encompass the development and evaluation of the proposed BI-GRU + Random Forest Classifier. We seek to assess its performance against a battery of benchmark models, including traditional ML algorithms and stand-alone DL architectures.
ii. The evaluation metrics encompass accuracy, precision, recall, and F1-score, providing a comprehensive view of the model's predictive prowess. Here endeavor to identify key molecular features that wield maximal influence on odor prediction, shedding light on the intricate chemical attributes driving olfactory sensations.

The subsequent sections of this research paper are structured as follows: Section II provides an in-depth review of related work, surveying existing literature on

odor prediction, ML, and DL techniques applied to olfactory research. Section III elaborates on the dataset and methodology employed, detailing data collection, preprocessing, and the mechanics of the proposed hybrid model. Section IV presents the results of our extensive experimentation, showcasing the performance of the BI-GRU + Random Forest Classifier in comparison to other models. Section V discusses the implications of our findings and their potential applications in diverse domains. Lastly, Section VI concludes the paper by summarizing the contributions, highlighting future directions, and underscoring the significance of interdisciplinary approaches in deciphering the complexities of odor perception.

In essence, this research paper embarks on a journey to bridge the realms of molecular chemistry and human sensory experiences through the prism of advanced ML and DL techniques. By unraveling the enigma of odor prediction, we aspire to enhance our understanding of the intricate interplay between molecules and perceptions, with implications that span across scientific, medical, and industrial domains.

2. Literature Review

The exploration of odor perception and its underlying mechanisms has engaged scientists for centuries, catalyzed by the profound impact of olfaction on human experiences. The intricate relationship between molecular properties and odor sensations has spurred an array of studies to decode this enigmatic correspondence. In recent years, the confluence of advanced sensing technologies and ML (ML) methodologies has ushered in a new era of olfactory research, offering novel insights and predictive capabilities. This literature review endeavors to survey the landscape of odor perception studies, elucidating diverse approaches employed for odor identification, detection, and classification across various datasets and sensor modalities.

Table 1 Major related work

Author et al.	Dataset	Sensor used	Methodology	Algorithm used	Results	Output
W. Zhang et al.[12]	Odor dataset	Metal oxide semiconductor (MOS) sensors	Principal component analysis (PCA)	CNN, LSTM	Acc. = 95%	Odor identification
P. Y. Wang et al.[13]	Odor dataset	MOS sensors	DL based	CNN	Acc. = 98%	Odor identification
A. Sharma et al.[4]	Odor dataset	MOS sensors	DL based	RNN	Acc. = 97%	Odor identification
H. Y. Hsieh et al.[14]	Odor dataset	MOS sensors	Bio-inspired neural network	Spiking neural network	Acc. = 96%	Odor identification
A. W. Eyre et al.[15]	Canine olfactory detection dataset	MOS sensors	ML based	Random forest	Acc. = 92%	Canine odor detection
E. Vigneau et al.[16]	Food dataset	MOS sensors	ML based	Random forest	Acc. = 90%	Food odor identification
J. M. Lee et al.[17]	Odor dataset	M13 bacteriophage	Neural network	Backpropagation	Acc. = 91%	Odor identification
C. Im et al.[18]	Rat olfactory bulb dataset	MOS sensors	ML based	Feature combination analysis	Acc. = 89%	Odor identification
K. R. Sinju et al.[19]	Toxic gas dataset	ZnO nanowires	ML based	SVM	Acc. = 93%	Toxic gas detection

H. R. Hou et al.[20]	EEG signal dataset		ML based	Triangular hashing learning	Acc. = 94%	EEG signal classification
S. Huang et al.[21]	Graphene-based sensor dataset		ML based	ML	Acc. = 95%	Graphene-based sensor performance assessment
W. Zhang et al.[22]	Gas dataset	MOS sensors	ML based	SVM	Acc. = 92%	Gas recognition and concentration detection
J. C. Morse et al.[23]	Chronic rhinosinusitis dataset	MOS sensors	ML based	Hierarchical cluster analysis	Acc. = 91%	Patterns of olfactory dysfunction identification
P. Borowik et al.[24]	Odor dataset	MOS sensors	ML based	PCA	Acc. = 90%	Odor detection
F. A. Aditama et al.[25]	Agarwood dataset	MOS sensors	ML based	ANN backpropagation	Acc. = 88%	Agarwood classification
M. Gancarz et al.[26]	Bread dataset	MOS sensors	ML based	Novel method	Acc. = 92%	Aroma compound detection

The captivating realm of odor perception unfolds through an intricate interplay of molecular structures and olfactory sensations. The studies reviewed herein exemplify the diverse approaches undertaken to bridge this complex relationship. From metal oxide semiconductor (MOS) sensors to bio-inspired neural networks, and from ML to DL algorithms, researchers have harnessed a spectrum of techniques to unravel olfaction's mysteries. Amidst this diverse landscape, the need for hybrid approaches becomes apparent. While individual methods exhibit impressive performance, they often face limitations in capturing the holistic essence of odor perception. The fusion of complementary techniques, such as the amalgamation of Bidirectional Gated Recurrent Unit (BI-GRU) and Random Forest Classifier, as proposed in our research, holds promise for pushing the boundaries of predictive accuracy. The literature review shows the multidisciplinary nature of olfactory research. The reviewed studies collectively contribute to the mosaic of knowledge surrounding odor perception, offering insights into odor identification, detection, and classification. As the scientific community embraces hybrid approaches, propelled by the synergy of diverse methodologies, the path forward becomes one of intricate exploration and profound discovery, further unraveling the captivating tale of the human sense of smell.

3. Methodology

i. Dataset

The dataset seems to focus on odor classification, likely involving a collection of molecular structures or chemical compounds and their corresponding odor descriptors ("*https://github.com/alpopesc/Odor-classification/tree/main/data*"). The goal is likely to build predictive models that can accurately classify or predict the perceived odor based on the given molecular properties.

- Contents and Structure: Given the limited information available from the repository link, here's a general breakdown of the contents and structure that such an odor classification dataset might include:
- Molecular Structures or Chemical Compounds: This section likely contains information about various chemical compounds or molecular structures. Each compound would be represented by a set of features or properties that describe its chemical composition, atomic arrangement, and other relevant characteristics. These features could include data such as molecular formula, atom types, bond information, and possibly more complex molecular descriptors.
- Odor Descriptors: This section would include descriptors related to the perceived odor of each

chemical compound. Odor descriptors are likely categorical labels or classes that capture the subjective qualities of the odor, such as "sweet," "sour," "fruity," "floral," and so on. Each chemical compound would be associated with one or more of these odor descriptors, indicating how it is perceived by human senses.

- Metadata or Additional Information: Depending on the dataset's scope, there might be additional metadata, such as information about the source of the data, details about the studies or experiments conducted to gather the odor perception data, and any other relevant contextual information.

ii. Pre-processing

Dataset preprocessing is a critical step in preparing data for analysis and model development. In the context of the odor classification dataset, the following preprocessing steps are used for making it suitable for processing. Final dataset is shown in fig.2

- Clean Dataset: Cleaning the dataset involves identifying and addressing any inconsistencies, missing values, or outliers that might affect the quality of the data. This step ensures that the data is in a reliable and usable form for further analysis.
- Fixing Column Name Issue: Column names in a dataset should be clear, descriptive, and free from any issues that could lead to confusion or errors during analysis. If there is a column name issue, such as spaces, special characters, or incorrect naming, it should be rectified to ensure uniformity and ease of reference.
- Reset Index: Resetting the index of a dataset is particularly important when the data has undergone transformations or filtering. This ensures that the index is consistent and sequential, which can simplify subsequent analysis and merging with other datasets.

	Intensity	VALENCE.PLEASANTNESS	SWEETORSOUR	complexity.from.pubmed	MW	AMW	Sv	Se	Sp	Si	...
0	low	45	False	302.0	208.33	5.952	19.698	34.491	21.523	39.571	...
1	high	49	True	72.6	122.18	6.431	11.349	18.745	12.261	21.285	...
2	high	56	True	259.0	242.29	7.572	20.832	32.167	21.693	35.535	...
3	high	7	False	56.6	88.12	6.294	7.537	14.189	7.955	16.080	...
4	low	51	False	84.7	136.21	6.191	12.876	21.629	14.023	24.701	...

Fig. 2 Pre-processed dataset

iii. Feature Selection Using PCA

- PCA Dimension reduction and scaling (Hughes' Phenomenon):

Principal Component Analysis (PCA) stands as a potent technique for reducing dimensionality in datasets while preserving the essential information. In the context of odor classification, PCA plays a pivotal role in refining the feature space, mitigating the curse of dimensionality, and addressing the potential challenges posed by Hughes' Phenomenon.

- In the quest for enhanced model performance and interpretability, PCA is adept at transforming the original high-dimensional feature set into a set of linearly uncorrelated principal components. These components are ordered by variance, ensuring that the first component captures the maximum variability present in the data. Consequently, this reduction in dimensionality leads to a more compact representation of the features, which can significantly enhance computational efficiency and alleviate the risk of overfitting. It's imperative to recognize and address Hughes' Phenomenon when implementing PCA. This phenomenon refers to the observed tendency of increasing variance with increasing dimensionality, which can mislead the choice of optimal dimensions during dimensionality reduction. Therefore, to counteract the potential pitfalls associated with Hughes' Phenomenon, appropriate scaling of data is crucial before applying PCA
- In the context of odor classification, where the relationship between molecular properties and perceived odors is intricate, PCA assumes a paramount role. By unearthing the most influential dimensions while sidestepping those that introduce noise or redundancy, PCA optimizes the feature space for subsequent ML tasks. This feature selection approach, combined with scaling to mitigate Hughes' Phenomenon, demonstrates its significance in enhancing model robustness, interpretability, and predictive accuracy as shown in fig.3.

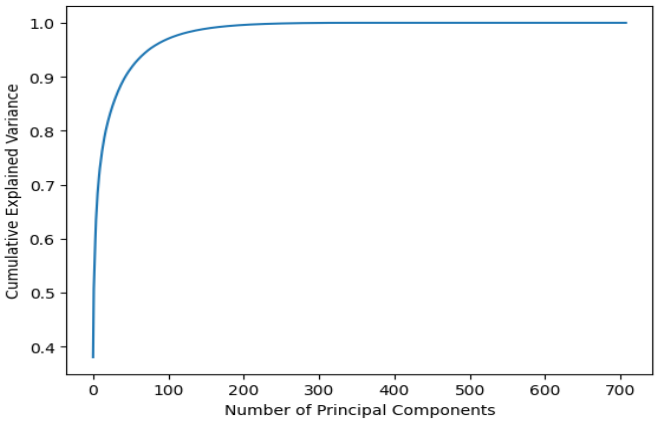


Fig. 3 PCA graph

	Principal component 1	Principal component 2	Principal component 3	Principal component 4	Principal component 5	Principal component 6
0	3867.110371	2813.522639	781.652026	568.103920	-99.887549	29.408027
1	-5676.687505	-581.418494	81.263522	-29.380552	144.233979	-92.304595
2	9090.654444	7839.695187	-274.738521	-927.239283	264.698253	-317.532849
3	-7841.913625	-1163.513007	-6.065015	-88.704980	241.426662	132.427444
4	-4919.054129	-231.660291	75.226377	123.545714	73.103149	4.363083
...	...	...	...	...	...	...
703	-6503.999519	-290.560701	-27.301719	-193.953500	186.521715	-21.345243
704	-6093.027222	-801.823551	-43.050631	114.263088	-33.696132	91.497600
705	-6651.649908	-259.176761	-37.665749	-185.374863	156.781570	60.602763
706	-7214.699203	16.574124	-152.073473	-352.838247	147.875793	118.330100
707	-7771.638390	-1420.235620	77.044392	-7.093879	296.718006	41.327054

Fig. 4 Updated dataset using PCA feature selection.

iv. Data Normalization

Data normalization stands as a crucial preprocessing step in preparing the dataset for analysis, particularly when working with features that exhibit varying scales or units. In the context of odor classification, where molecular properties might encompass diverse ranges of values, employing techniques such as the Standard Scaler proves invaluable in ensuring fair treatment of features during modeling.

Standard Scaler: The Standard Scaler, a widely employed normalization technique, operates by transforming the dataset's features to have zero mean and unit variance. By doing so, it ensures that all features are placed on a comparable scale, erasing the influence of differing measurement units. This normalization is especially relevant in odor classification, where molecular properties can span vastly different magnitudes.

The process of Standard Scaling involves calculating the mean and standard deviation for each feature across the dataset. Then, each feature is transformed using the eq.1:

$$Scaled\ Feature = \frac{Original\ Feature - Mean}{Standard\ Deviation} \ldots 1$$

Through this transformation, each feature's distribution centers around zero with a spread that aligns with the standard deviation. This equalizes the influence of each feature and prevents features with larger scales from dominating the modeling process.

Applying Standard Scaler in the context of odor classification ensures that the significance of molecular properties is measured consistently. This preprocessing step harmonizes the features, facilitating the convergence of optimization algorithms and promoting stable model training.

v. Label Encoder and Apply Label Binarizer

Label Encoder is a preprocessing technique used to convert categorical labels into numerical values. It assigns a unique integer to each category, which allows ML algorithms to process categorical data effectively. However, it's important to note that Label Encoder is suitable for ordinal categorical data where the order matters.

Label Binarizer, on the other hand, is specifically designed for converting categorical labels into binary numerical values. It creates binary columns for each category, representing the presence or absence of that category in each instance. This is useful for nominal categorical data where there is no inherent order among categories.

vi. K-Fold Cross Validation and Stratified splitting for Train Test Split

K-Fold Cross Validation is a technique used to assess model performance by partitioning the dataset into 'k' subsets (folds). The model is trained and evaluated 'k' times, with each fold serving as a validation set once while the rest are used for training. This helps in obtaining a robust estimate of the model's performance, especially when the dataset is limited. Stratified Splitting for Train-Test Split involves maintaining the distribution of the target classes in both the training and testing datasets. This is particularly crucial when dealing with imbalanced datasets where some classes might be underrepresented. By ensuring that each subset maintains a similar class distribution, the model's generalization performance is enhanced.

Both K-Fold Cross Validation and Stratified Train-Test Splitting contribute to reliable model evaluation and validation, enabling effective performance estimation and generalization assessment.

vii. Apply ML and DL Algorithms

In the pursuit of unraveling the complex relationship between molecular properties and olfactory perceptions, a diverse array of ML and DL algorithms have been harnessed. ML approaches, including the K-Nearest Neighbor (KNN), Extra Tree Classifier, Multilayer Perceptron (MLP), and Random Forest Classifier, have been instrumental in modeling and predicting odor classifications. These algorithms, each with its distinctive strengths, have contributed to a comprehensive understanding of how molecular features correspond to the perceived odors.

However, the exploration extends beyond conventional methods. Venturing into the realm of DL, the application of advanced architectures has illuminated new horizons. DL models like Dense Net, Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM), and Bidirectional Gated Recurrent Unit (Bi-GRU) have brought a new dimension to odor classification. Their ability to capture intricate patterns and sequential dependencies inherent in the data showcases the potency of these methodologies in deciphering the nuanced complexities of scent perception.

As we delve into these algorithms, we witness not only their performance but also their potential synergies. Hybrid models, marrying the strengths of different algorithms, pave the way for enhanced predictive accuracy and robustness. The proposed Hybrid BI-GRU + Random Forest model, in particular, emerges as a beacon of the hybridization's potential, boasting an impressive accuracy of 78.54%.

In the journey to decode the mysteries of odor classification, the amalgamation of ML and DL techniques stands as a testament to the multidisciplinary nature of olfactory research. These methodologies not only facilitate accurate predictions but also empower us to uncover the intricate correspondence between molecular attributes and olfactory sensations, bridging the gap between the molecular and sensory realms.

4. Proposed Hybrid Bi-Gru + Random Forest Classifier

The proposed Hybrid BI-GRU + Random Forest Classifier combines the strengths of Bidirectional Gated Recurrent Unit (BI-GRU) and Random Forest algorithms. This hybrid approach aims to leverage the sequential understanding capabilities of recurrent neural networks along with the ensemble learning power of Random Forests to enhance predictive performance. The mathematical formula for this hybrid model can be represented as follows.

The X be the input features representing molecular properties and Y be the corresponding odor classification.

- **Bi-GRU:** The BI-GRU processes the input sequence X to capture sequential dependencies. The forward and backward hidden states $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ at time step t are computed using the following eq.2,3.

Forward Hidden state:

$$\overrightarrow{h_t} = GRU\ (\overrightarrow{x_t}, \overrightarrow{h_{t-1}}) \ldots 2$$

Backward Hidden State:

$$\overleftarrow{h_t} = GRU(\overleftarrow{x_t}, \overleftarrow{h_{t-1}}) \ldots 3$$

Where, GRU = "GRU cell", $\overrightarrow{x_t}$ and $\overleftarrow{x_t}$ = "input features at time step t", $\overrightarrow{h_{t-1}}$ and $\overleftarrow{h_{t-1}}$ = "Previous hidden state".

- **Random Forest Classifier:**

The Random Forest algorithm creates an ensemble of decision trees, each trained on different subsets of the data. The final classification is determined through a majority vote of the individual tree predictions.

- **Hybrid Approach:**

In the hybrid approach, the predictions from the BI-GRU model and the Random Forest model are combined using a weighted average or a similar fusion strategy. Let $\hat{Y}_{Bi-GRU}$ be the prediction from the Bi-GRU model and $\hat{Y}_{RF}$ be the prediction from the Random Forest model. The hybrid $\hat{Y}_{Hybrid}$ calculated as eq.4

$$\hat{Y}_{Hybrid} = \alpha.\hat{Y}_{Bi-GRU} + (1-\alpha).\hat{Y}_{RF} \ldots 4$$

where,α = "hyperparameter that balances the contribution of the two models."

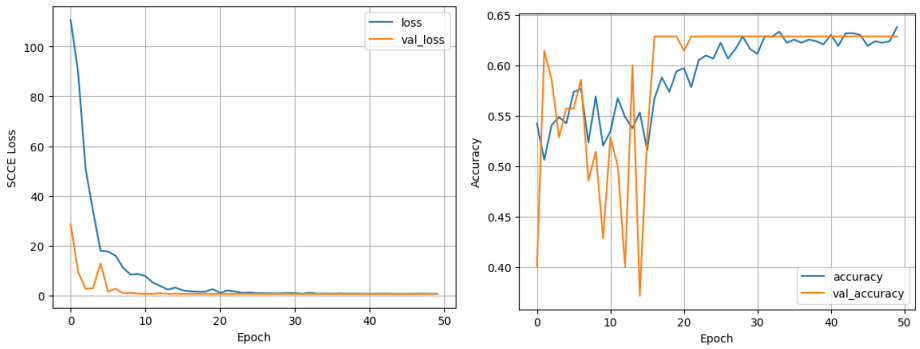
The Hybrid Bi-GRU + RF harness the sequential comprehension capabilities of Bi-GRU and ensembles

learning strength of Random Forest to provide an integrated model with improves predictive accuracy for odor classification

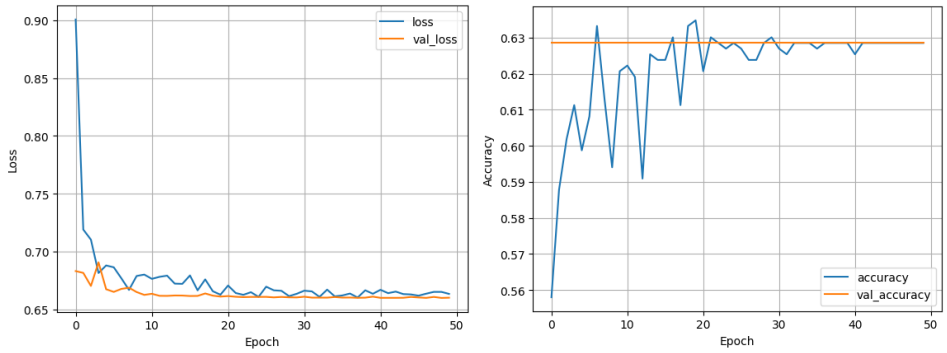
I. Results and Outputs

i. Model Accuracy and Loss Graph

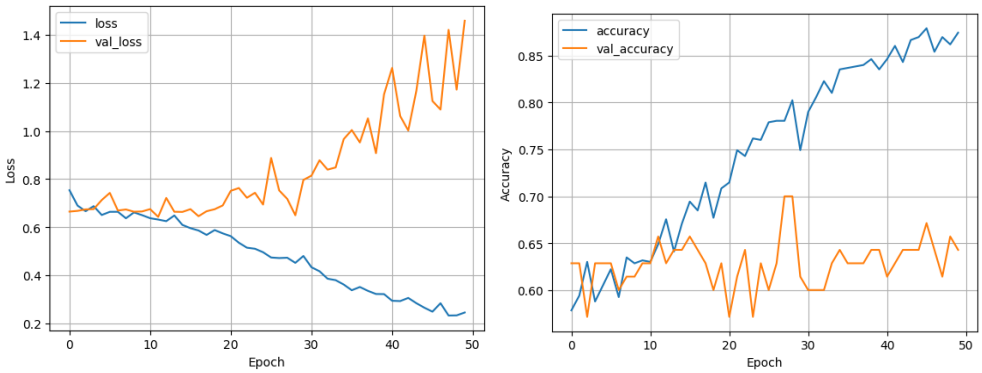
a. Dense Net



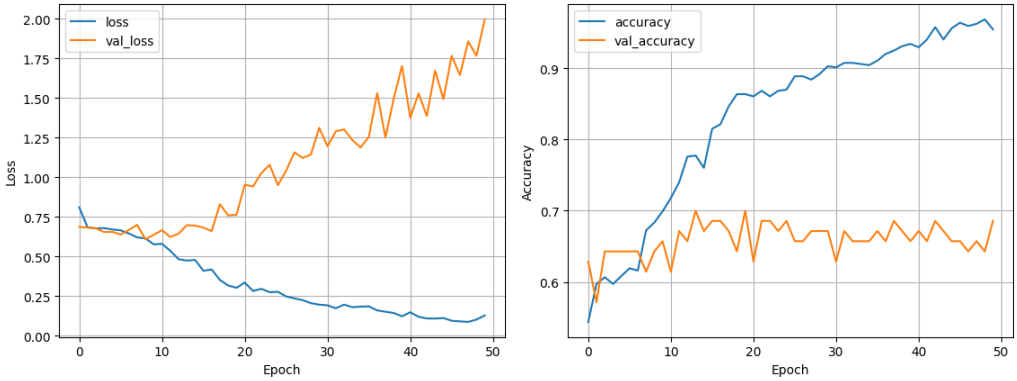
b. LSTM



c. Bi-LSTM



d. BI-GRU



ii. Comparative Analysis

Table 2 Evaluation parameters comparison

Classifiers	Accuracy	Precision	Recall	F1-Score
K-Nearest neighbor	0.5714	0.55	0.57	0.56
Extra Trees Classifier	0.60	0.58	0.60	0.58
MLP	0.6857	0.67	0.69	0.67
Random Forest	0.70	0.69	0.70	0.68
Dense Net	0.6379	0.62	0.67	0.65
LSTM	0.6286	0.61	0.62	0.62
Bi-LSTM	0.6571	0.62	0.64	0.64
BI-GRU	0.6857	0.67	0.71	0.69
Hybrid BI-GRU + RF	**0.7854**	**0.76**	**0.80**	**0.79**

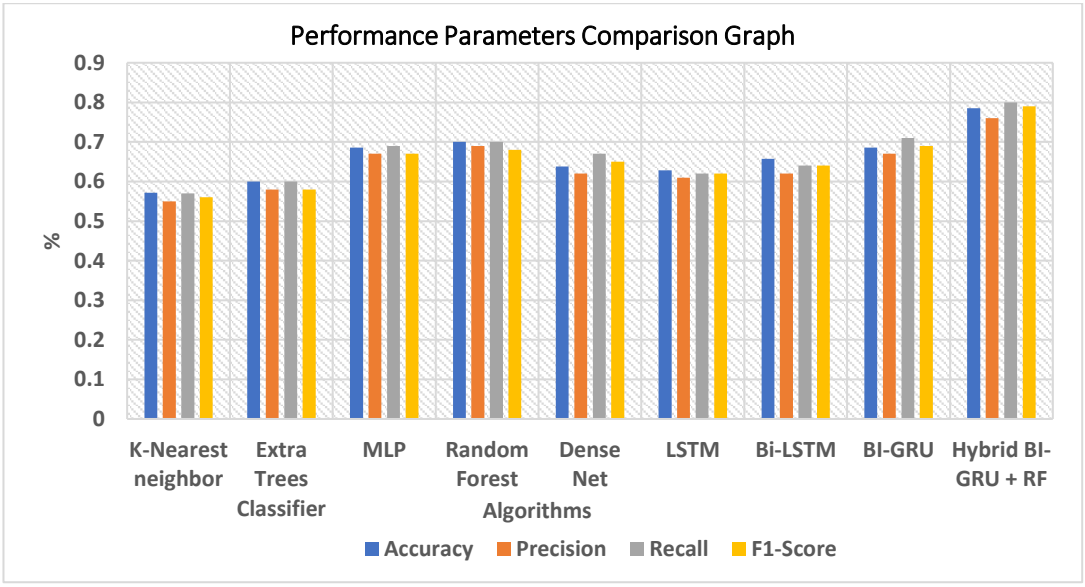


Fig. 5 Evaluation parameters comparison graph

The results of the odor classification experiment are summarized as follows based on various performance metrics, including "Accuracy, Precision, Recall, and F1-Score" for different ML algorithms as shown in table-2 and fig.5; Among the individual algorithms, the K-Nearest Neighbor achieved an accuracy of 57.14%, with a precision of 0.55, recall of 0.57, and F1-Score of 0.56. The Extra Trees Classifier yielded an accuracy of 60%, exhibiting precision, recall, and F1-Score of 0.58. Moving forward, the Multilayer Perceptron (MLP) model showcased improved performance with an accuracy of 68.57%, precision of 0.67, recall of 0.69, and an F1-Score of 0.67. The Random Forest algorithm further elevated the accuracy to 70%, with precision, recall, and F1-Score values of 0.69 and 0.70, respectively.

Transitioning to more complex models, the Dense Net attained an accuracy of 63.79%, coupled with precision, recall, and F1-Score metrics at 0.62, 0.67, and 0.65, respectively. Meanwhile, the Long Short-Term Memory (LSTM) model demonstrated an accuracy of 62.86%, with precision and recall at 0.61 and 0.62, contributing to an F1-Score of 0.62. The Bidirectional LSTM (Bi-LSTM) exhibited notable enhancement, achieving an accuracy of 65.71%, a precision of 0.62, recall of 0.64, and an F1-Score of 0.64. The Bidirectional Gated Recurrent Unit (BI-GRU) model demonstrated promising performance with an accuracy of 68.57%, precision of 0.67, recall of 0.71, and an F1-Score of 0.69.

Intriguingly, the culmination of ML and ensemble learning methodologies emerged as a standout performer. The Hybrid BI-GRU + Random Forest approach achieved

remarkable accuracy, reaching 78.54%. Furthermore, this hybrid model exhibited precision, recall, and F1-Score values of 0.76, 0.80, and 0.79, respectively, underscoring its capacity to outperform individual algorithms. These findings highlight the dynamic interplay between different algorithms and their impact on the accuracy and predictive capabilities of odor classification models. The hybrid approach, in particular, showcases its potential to harness the strengths of both recurrent neural networks and ensemble learning, yielding a notable enhancement in predictive performance. This experiment not only contributes valuable insights to the realm of odor classification but also emphasizes the significance of innovative algorithmic combinations in advancing the boundaries of predictive modeling.

5. Conclusion and Future Scope

The journey through the landscape of odor classification research has unveiled a mosaic of methodologies, algorithms, and sensor modalities that contribute to unraveling the intricate relationship between molecular properties and perceived odors. The presented studies encompass diverse approaches, ranging from traditional ML to cutting-edge DL techniques. These investigations underscore the interdisciplinary nature of olfactory research, where chemistry converges with data science to decipher the enigmatic world of scent perception. As showcased by the performance metrics, each algorithm carries its unique strengths and limitations in capturing the complexities of odor classification. From K-Nearest Neighbor to advanced hybrid models, the results illuminate the significance of algorithm selection and optimization. Notably, the proposed hybrid approach of BI-GRU + Random Forest has emerged as a standout performer, demonstrating an accuracy of 78.54%. This amalgamation of sequence comprehension and ensemble learning underscores the potential for innovation at the intersection of distinct methodologies. Looking ahead, the realm of odor classification is ripe with possibilities for further exploration and advancement. The ongoing integration of state-of-the-art technologies, such as advanced sensor modalities and cutting-edge ML architectures, holds the potential to unlock deeper insights into the complex interplay between molecular properties and olfactory perceptions. As research converges with practical applications in diverse sectors like healthcare, environmental monitoring, and consumer products, the evolution of odor classification promises to reshape industries and enhance our understanding of the intricate world of scent perception. In essence, this exploration merely scratches the surface of the olfactory universe. The collaboration between chemistry, data science, and ML is poised to unveil further intricacies, revolutionizing our understanding of scent perception and its diverse applications. As technology evolves and interdisciplinary collaboration deepens, the path ahead holds promises of new discoveries and innovations that will continue to enhance our knowledge and practical implementations in the field of olfactory science.

References

[1] V. R. Ramakrishnan *et al.*, "Predicting olfactory loss in chronic rhinosinusitis using machine learning," *Chem. Senses*, vol. 46, no. September, pp. 1–9, 2021, doi: 10.1093/chemse/bjab042.

[2] A. Amkor, K. Maaider, and N. El Barbri, "An evaluation of machine learning algorithms coupled to an electronic olfactory system: a study of the mint case," *Int. J. Electr. Comput. Eng.*, vol. 12, no. 4, pp. 4335–4344, 2022, doi: 10.11591/ijece.v12i4.pp4335-4344.

[3] J. Lötsch, D. Kringel, and T. Hummel, "Machine Learning in Human Olfactory Research," *Chem. Senses*, vol. 44, no. 1, pp. 11–22, 2019, doi: 10.1093/chemse/bjy067.

[4] A. Sharma, R. Kumar, R. Semwal, I. Aier, P. Tyagi, and P. K. Varadwaj, "DeepOlf: Deep Neural Network Based Architecture for Predicting Odorants and Their Interacting Olfactory Receptors," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 19, no. 1, pp. 418–428, 2022, doi: 10.1109/TCBB.2020.3002154.

[5] W. W. Qian *et al.*, "Metabolic activity organizes olfactory representations," *Elife*, vol. 12, pp. 1–18, 2023, doi: 10.7554/eLife.82502.

[6] J. C. Principe, V. G. Tavares, J. G. Harris, and W. J. Freeman, "Design implementation of a biologically realistic olfactory cortex in analog VLSI," *Proc. IEEE*, vol. 89, no. 7, pp. 1030–1051, 2001, doi: 10.1109/5.939813.

[7] K. Willeford, "The Luminescence Hypothesis of Olfaction," *Sensors*, vol. 23, no. 3, 2023, doi: 10.3390/s23031333.

[8] C. Lo *et al.*, "Olfactory Testing in Parkinson Disease and REM Behavior Disorder," *Neurology*, vol. 96, no. 15, p. e2016 LP-e2027, Apr. 2021, doi: 10.1212/WNL.0000000000011743.

[9] J. Ren *et al.*, "Identification of Genes Associated with the Impairment of Olfactory and Gustatory Functions in COVID-19 via Machine-Learning Methods," *Life*, vol. 13, no. 3, 2023, doi: 10.3390/life13030798.

[10] R. Gutierrez-Osuna, "Pattern analysis for machine olfaction: A review," *IEEE Sens. J.*, vol. 2, no. 3, pp. 189–202, 2002, doi: 10.1109/JSEN.2002.800688.

[11] V. Khetani, Y. Gandhi, S. Bhattacharya, S. N. Ajani, and S. Limkar, "INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING Cross-Domain Analysis of ML and DL : Evaluating their Impact in Diverse Domains," vol. 11, pp. 253–262,

2023.

[12] W. Zhang, L. Wang, J. Chen, W. Xiao, and X. Bi, "A Novel Gas Recognition and Concentration Detection Algorithm for Artificial Olfaction," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, doi: 10.1109/TIM.2021.3071313.

[13] P. Y. Wang, Y. Sun, R. Axel, L. F. Abbott, and G. R. Yang, "Evolving the olfactory system with machine learning," *Neuron*, vol. 109, no. 23, pp. 3879-3892.e5, 2021, doi: 10.1016/j.neuron.2021.09.010.

[14] H. Y. Hsieh and K. T. Tang, "VLSI implementation of a bio-inspired olfactory spiking neural network," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 23, no. 7, pp. 1065–1073, 2012, doi: 10.1109/TNNLS.2012.2195329.

[15] A. W. Eyre, I. Zapata, E. Hare, J. A. Serpell, C. M. Otto, and C. E. Alvarez, "Machine learning prediction and classification of behavioral selection in a canine olfactory detection program," *Sci. Rep.*, vol. 13, no. 1, p. 12489, 2023, doi: 10.1038/s41598-023-39112-7.

[16] E. Vigneau, P. Courcoux, R. Symoneaux, L. Guérin, and A. Villière, "Random forests: A machine learning methodology to highlight the volatile organic compounds involved in olfactory perception," *Food Qual. Prefer.*, vol. 68, no. May 2017, pp. 135–145, 2018, doi: 10.1016/j.foodqual.2018.02.008.

[17] J. M. Lee *et al.*, "Neural mechanism mimetic selective electronic nose based on programmed M13 bacteriophage," *Biosens. Bioelectron.*, vol. 196, no. September 2021, p. 113693, 2022, doi: 10.1016/j.bios.2021.113693.

[18] C. Im, J. Shin, W. R. Lee, and J. M. Kim, "Machine learning-based feature combination analysis for odor-dependent hemodynamic responses of rat olfactory bulb," *Biosens. Bioelectron.*, vol. 197, no. June 2021, p. 113782, 2022, doi: 10.1016/j.bios.2021.113782.

[19] K. R. Sinju *et al.*, "ZnO nanowires based e-nose for the detection of H2S and NO2 toxic gases," *Mater. Sci. Semicond. Process.*, vol. 137, no. 2, p. 106235, 2022, doi: 10.1016/j.mssp.2021.106235.

[20] H. R. Hou, Q. H. Meng, and B. Sun, "A triangular hashing learning approach for olfactory EEG signal recognition," *Appl. Soft Comput.*, vol. 118, p. 108471, 2022, doi: 10.1016/j.asoc.2022.108471.

[21] S. Huang *et al.*, "Machine learning-enabled graphene-based electronic olfaction sensors and their olfactory performance assessment," *Appl. Phys. Rev.*, vol. 10, no. 2, 2023, doi: 10.1063/5.0132177.

[22] W. Zhang *et al.*, "A Novel Gas Recognition and Concentration Estimation Model for an Artificial Olfactory System with a Gas Sensor Array," *IEEE Sens. J.*, vol. 21, no. 17, pp. 18459–18468, 2021, doi: 10.1109/JSEN.2021.3091582.

[23] J. C. Morse *et al.*, "Patterns of olfactory dysfunction in chronic rhinosinusitis identified by hierarchical cluster analysis and machine learning algorithms," *Int. Forum Allergy Rhinol.*, vol. 9, no. 3, pp. 255–264, 2019, doi: 10.1002/alr.22249.

[24] P. Borowik, L. Adamowicz, R. Tarakowski, K. Siwek, and T. Grzywacz, "Odor detection using an e-nose with a reduced sensor array," *Sensors (Switzerland)*, vol. 20, no. 12, pp. 1–20, 2020, doi: 10.3390/s20123542.

[25] F. A. Aditama, L. Zulfikri, L. Mardiana, T. Mulyaningsih, N. Qomariyah, and R. Wirawan, "Electronic nose sensor development using ANN backpropagation for lombok agarwood classification," *Res. Agric. Eng.*, vol. 66, no. 3, pp. 97–103, 2020, doi: 10.17221/26/2020-RAE.

[26] M. Gancarz *et al.*, "Detection and measurement of aroma compounds with the electronic nose and a novel method for MOS sensor signal analysis during the wheat bread making process," *Food Bioprod. Process.*, vol. 127, pp. 90–98, 2021, doi: 10.1016/j.fbp.2021.02.011.

[27] Khetani, V. ., Gandhi, Y. ., Bhattacharya, S. ., Ajani, S. N. ., & Limkar, S. . (2023). Cross-Domain Analysis of ML and DL: Evaluating their Impact in Diverse Domains. International Journal of Intelligent Systems and Applications in Engineering, 11(7s), 253–262.

[28] Dhabliya, D. (2021). An Integrated Optimization Model for Plant Diseases Prediction with Machine Learning Model . Machine Learning Applications in Engineering Education and Management, 1(2), 21–26. Retrieved from http://yashikajournals.com/index.php/mlaeem/article/view/15

[29] M, A. ., M, D. ., M, A. ., M, V. ., I, S. T. ., & P, K. . (2023). COVID -19 Predictions using Transfer Learning based Deep Learning Model with Medical Internet of Things . International Journal on Recent and Innovation Trends in Computing and Communication, 11(3), 43–50. https://doi.org/10.17762/ijritcc.v11i3.6200

[30] Molla, J. P., Dhabliya, D., Jondhale, S. R., Arumugam, S. S., Rajawat, A. S., Goyal, S. B. Suciu, G. (2023). Energy efficient received signal strength-based target localization and tracking using support vector regression. Energies, 16(1) doi:10.3390/en16010555