Honey Adulteration Detection via Robust Diffusion Classifier and Hyperspectral Imaging

Weihao Tang
The University of Auckland
Auckland, New Zealand
wtan402@aucklanduni.ac.nz

Guyang Zhang
The University of Auckland
Auckland, New Zealand
gzha422@aucklanduni.ac.nz

Waleed Abdulla
The University of Auckland
Auckland, New Zealand
w.abdulla@auckland.ac.nz

***Abstract*—Honey is the third-most adulterated food product in the world. Recently, AI-based hyperspectral detection has exhibited promising superiority in classifying pure and adulterated honey. However, the classification robustness of hyperspectral-based models has barely been assessed in existing studies, leading to the classification model being catastrophically "collapsed" under the attack of perturbations from honey adulteration. A trustworthy detective model is anticipated to classify the adulterated honey reliably, irrespective of perturbation attacks. In light of this, we propose a modeling strategy that views adulteration perturbations as variables. On top of that, accumulated negative log-likelihoods of conditional generative models are further wrapped with the Bayesian rule to certify the classification results, which improves classification robustness and accuracy of spectral detection models under attacks of adulteration perturbations. The results of empirical experiments imply that our proposed method successfully outperforms the state-of-the-art in applying hyperspectral techniques to classify the authenticity of honey. The precision score is still averagely over 85% for classifying adulterated honey when adulteration concentrations are shifted, and the recall score is over 85% when unobserved honey types attack. Consequently, our findings are eligible to serve as a potential way to improve the robustness and accuracy of hyperspectral detection models against food fraud.**

***Index Terms*—Manuka honey, Hyperspectral imaging, Diffusion models, Syrup adulteration.**

I. Introduction

The global honey market generated a revenue of USD 9,448.1 million in 2023 and is expected to reach USD 13,574.7 million by 2030. Unfortunately, honey is the third most adulterated food product in the world [1]. Honey adulteration will pose severe risks to the glycemia and diet of consumers, as the addition of cheap syrup is the primary adulteration trick [2]. Hence, it is significant to classify authentic and adulterated honey reliably to protect consumers' health and wealth. Recently, AI-based spectral detection has exhibited elite superiority in classifying the authenticity of honey, irrespective of cost, efficiency, and non-destructiveness [3]–[5].

However, most studies have not fully discussed the classification robustness of honey detection models, resulting in the penalty of models poorly classifying the spectrum of adulterated honey when different adulterated tricks produce some shifts. In this scenario, our detection model will be fooled to flip its prediction if spectral inputs are perturbed with unobserved adulteration shifts from syrup-honey concentrations. In order to address this issue, a straightforward solution is to manually simulate as many adulterated samples with shifted syrup-honey ratios as possible. Subsequently, implementing adversarial training improves the robustness of classification models against attacks of adulteration perturbation [6], [7]. In actuality, artificially mixing syrup and honey via an expected ratio is too labor-intensive, time-consuming, and complicated.

Currently, the reference solution is to generate perturbed samples through a generative model or augment perturbed samples [7], [8] rather than producing them artificially. Further, a larger scale of training on top of the expanded database is to improve the robustness and generalization of the classifier [9]. Although this two-step approach achieves the state-of-the-art in detecting honey quality, data generation and classification processes are independent; thereby, it fails to link the generative models' simulated trajectories [10] with class predictions. In terms of this, we manage to bridge two separable steps by accumulating full simulated path $p_t^\theta(\mathbf{x}|c)$ to estimate negative log-likelihoods (NLL) for classification [11]. In light of the estimated NLLs under different class conditions, Bayesian inference is conducted to recognize honey authenticity despite the uncertainty of perturbation robustly.

In conclusion, we intend to handle the underlying adulterated attacks on classification models, such as human-designed shifts in syrup-honey ratios. As illustrated in Fig. I, a conditional generative model is implemented to further improve the spectral detection model's robustness and accuracy in classifying honey's authenticity. Technically, the conditional diffusion model (CDM) [12] is conducted to offer a certified robust classification by using approximated ELBOs to calculate the maximum posterior. The results demonstrate that our proposed method outperforms conventional methods in honey spectral detection. In particular, we observe exceptional performance in terms of the model's robustness against adulteration perturbations. In summary, our study makes three main contributions:

(1) Firstly, we propose a novel causal map, viewing the adulteration perturbation, such as shifted syrup-honey ratios and unobserved honey types, as variables during hyperspectral authenticating.

(2) On top of this map, the Bayesian rule is introduced to calculate the classification probability by estimating the negative log-likelihoods of conditional flows.

(3) Finally, the probabilistic model allows for robust classification of authentic and adulterated honey even when it suffers

979-8-3315-7206-8/25/$31.00

from adulteration perturbations.

II. Related work

This section begins with current works establishing a hyperspectral classification model against food fraud. Subsequently, we will concentrate on recent advances in conducting generative models to detect food fraud.

A. Spectral classification model against honey fraud

Currently, the construction of hyperspectral classification models for honey adulteration mainly depends on two training methods: statistical machine learning-based [3] and deep learning-based [13]. Although the former has the advantages of cost-free training and theoretical evidence, poor representation learning limits its performance due to the high dimensionality and complexity of hyperspectral data [14]. As a result, owing to the powerful representation learning ability, deep learning has successfully been a mainstream evolution in detecting honey adulteration by hyperspectral techniques [2]. However, supervised deep learning has an extensive demand for the training data scale [15]. Otherwise, the classification model will suffer from catastrophic uncertainty, especially when input spectral data is perturbed. To ensure a robust classification model under attacks of adulteration, manually simulating thousands of honey-adulterated samples with diverse syrup-honey ratios might be mandatory. However, obtaining such a highly anticipated database is too complex, thus forcing us to explore more reasonable solutions to improve the classification robustness of hyperspectral detection models against honey fraud.

B. Conditional diffusion models

Recently, there have been rapid findings that generative models proceed with areas such as image-audio generation, 3D splatting, and image segmentation [16]. Notably, generative models have been shown to be feasible in facilitating food fraud detection. GANs and CVAEs are the two most popular models used to generate additional spectrum and attack the spectral detection model [6], [7]. Despite the certified advantages of GANs and CVAEs in generating the training spectrum with our target conditions, recent work implies that diffusion models beat them in multiple dimensions [10]. In particular, diffusion or score models are more promising in dealing with uncertainty, which is the primary focus of our study, since they simulate a time-dependent stochastic process where data are gradually corrupted with noise and reconstructed in a probabilistic way [10].

CDM represents a powerful and flexible class of generative models that extend the capabilities of score-based diffusion processes to handle conditional data generation tasks. Based on the foundation of stochastic differential equations (SDEs), these models learn to reverse a predefined diffusion process that gradually transforms structured data into noise. By conditioning on auxiliary information, such as class labels, text, or other modalities, conditional diffusion models can generate samples that are not only high-quality but also aligned with the provided conditioning signal.

C. Lipshcitz constant of classifiers

In the view of a classifier $\{_\theta(\mathbf{x}) \in \mathbb{R}^D$, which represents input $\mathbf{x}$ to output logits or class probabilities over the class y. To measure the smoothness or robustness of the classifier in the presence of noise perturbations, the definition of the Lipschitz constant [17] is in local Lipschitz continuity:

$$\|f_\theta(\mathbf{x}+\boldsymbol{\epsilon}) - f_\theta(\mathbf{x})\| \leq K_0\|\epsilon\| \tag{1}$$

Alternatively, we can simplify the Lipschitz constant of a deep learning classifier via its first-order gradient of ϵ in Eq. 2. The format of this constant could be updated if we have a smoothed function [18].

$$K_0 = \sup_{\mathbf{x}} \|\nabla f(\mathbf{x})\| = \max_{\mathbf{u}} \mathbf{u}^\top \nabla \mathbb{E}_\epsilon \left[f(\mathbf{x}+\epsilon)\right] \tag{2}$$

Here, K_0 is the Lipschitz constant and $\mathbb{E}_\epsilon[.]$ is a smoothed continuous function. The property K_0 ensures that small changes in the input of the function result in proportionally small changes in the output, which is crucial in many areas like numerical analysis and machine learning, particularly in the training and convergence of models.

III. Proposed method

This section first introduces our proposed causal map for modeling the process of honey authentication. Secondly, we implement CDMs to improve classification robustness.

A. Proposed causal map of modeling strategy

Fig. III-A compares the previous strategy and ours. One of the most evident differences is that existing works barely treat those shifts in adulteration concentrations as variables. Instead, only a few separated and fixed values were prepared to train a spectral detective model. However, this training approach will penalize the model's robustness against perturbations, regardless of any variations and shifts in the detection process. Particularly, perturbed samples in our study refer to spectral data outside the distribution constructed during the learning process due to artificially designed shifts in adulterated ratios. As far as we are concerned, this will be the first work to evolve the traditional discriminative network with a generative network in food adulteration detection. Owing to this advantage, we intend to utilize denoising and stochastic processes modeling from diffusion models to purify the perturbation and resolve the uncertainties caused by out-of-distribution (OOD) spectral data by shifted syrup-honey ratios.

B. Robust classifier via conditional diffusion models

Score-based diffusion models have recently become the most attractive topic in tasks like image generation and audio synthesis [10]. It refers to a family that involves gradually diffusing the data distribution towards a given noise distribution using a stochastic differential equation in Eq. 3 and learning the time reversal of this SDE for sample generation. Apart from the CVAEs or GANs, the diffusion process starts from the class-condition data $\mathbf{x}_0 \sim p(\mathbf{x}_0|c)$ and moves to the noise data $\mathbf{x}_T \sim \mathcal{N}(0, I)$; meanwhile, the time evolution path of

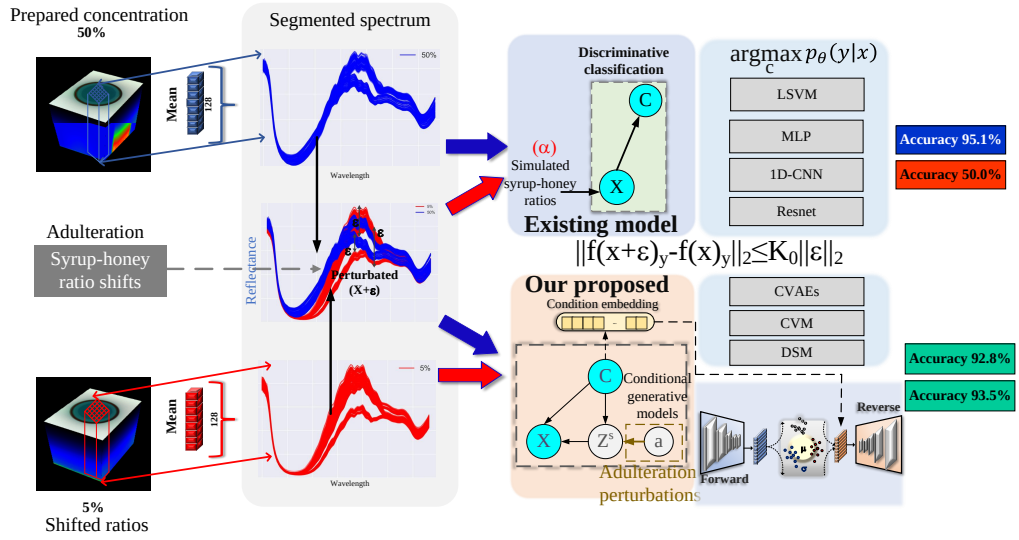

Fig. 1. A modeling strategy for a robust detection model of adulterated honey in the presence of unknown ratios is proposed.

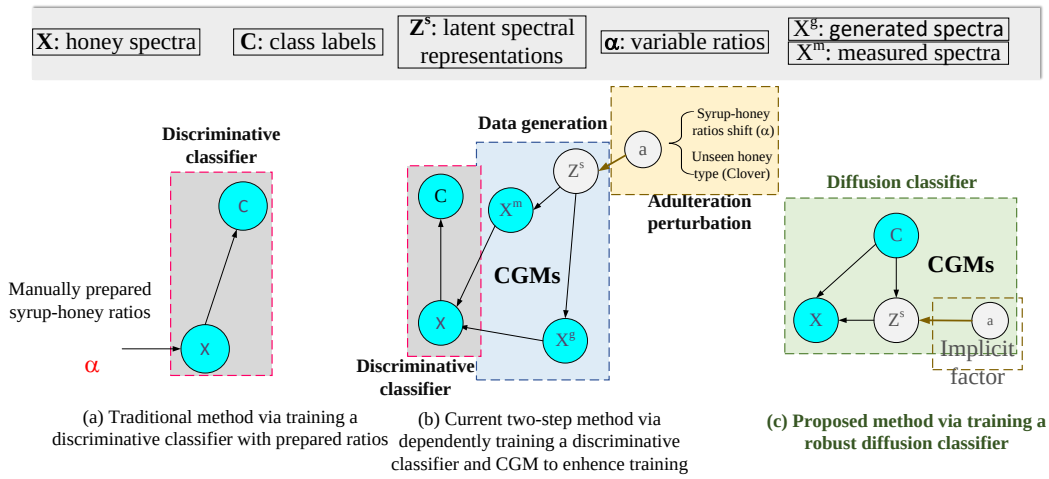

Fig. 2. Proposed modeling strategy of detecting honey authenticity by using diffusion classifiers.

$p_t(\mathbf{x})$ is governed by the Fokker-Planck equation denoted as Eq.4. Accordingly, the training loss of our diffusion model could be written as Eq. 5.

$$\mathrm{d}\mathbf{x}_t = f_t(\mathbf{x}_t)\,\mathrm{d}t + g_t\,\mathrm{d}\mathbf{w}_t \tag{3}$$

$$\frac{\partial p_t(\mathbf{x})}{\partial t} = -\nabla_{\mathbf{x}} \cdot (\mathbf{f}(\mathbf{x},t)p_t(\mathbf{x})) + \frac{1}{2}\Delta(g^2(t)p_t(\mathbf{x})) \tag{4}$$

$$\begin{aligned}&\mathcal{J}_{DSM}^{cond}(\theta;g(t),c)\\ &\quad = \mathbb{E}_{\mathbf{x},t}[g(t)^2\|\mathbf{s}_\theta(\mathbf{x}_t,t) - \nabla_{\mathbf{x}'}\log p_t(\mathbf{x}'|\mathbf{x})\|_2^2]\end{aligned} \tag{5}$$

Song et al. [10] revealed that with a designed weighting, the score-matching losses upper bound the log-likelihood of score-based models. According to this theory, we approximate the negative log-likelihood of score models with its upper bound, denoted as Eq. 6. To calculate the classification probability via the Bayesian rule, approximated classification ELBOs of CDM are implemented to obtain the posterior through Eq. 8. The maximum posterior determines the final output class. The completed workflow is detailed in Algorithm 1. Additionally, the Lipschitz constant of a classifier implies the robustness and smoothness ability to go against the input noise $\mathbf{x} + \epsilon$. CDM-based classifier's robustness will be certified according to its Lipschitiz continuity assessment in Section IV-C.

$$-\log p_\theta^{SDE}(\mathbf{x}_{\text{data}}|c) \leq \mathcal{L}_\theta^{CDM}(\mathbf{x};c) \tag{6}$$

$$\begin{aligned}&\mathcal{L}_{\boldsymbol{\theta}}^{CDM}(\mathbf{x};c)\\ &\quad = \mathbb{E}_{\mathbf{x},t}[g(t)^2\|\mathbf{s}_\theta(\mathbf{x}_t,t) - \nabla_{\mathbf{x}'}\log p_t(\mathbf{x}'|\mathbf{x})\|_2^2] + C\end{aligned} \tag{7}$$

$$p_\theta(c|\mathbf{x}) = \textbf{softmax}(\log p_\theta(\mathbf{x}|c)) \tag{8}$$

Algorithm 1: Robust classification via CDSM

Input: A pre-trained score model $\mathbf{s}_\theta$, input perturbed spectral data $\mathbf{x}'$, diffusion steps T, optimization steps N, number of all class conditions K, importance weighting $\lambda(t)$, drift term $f(\mathbf{x_t}, t)$, diffusion term $g(t)$.

Output: Predicted class label $\tilde{c}$

Result: Class prediction $\tilde{c} = \arg\max_c p_\theta(c|x)$

Initialize : $m = 0, \hat{\mathbf{x}} = \mathbf{x}', \lambda(t) \leftarrow g(t)^2, \tilde{f}_{DSM}(0) \leftarrow 0$

for $n \in \{1, 2, \ldots, N-1\}$ **do**

  Estimate $\mathrm{g} = \nabla_{\hat{\mathbf{x}}} \mathbb{E}_{\hat{\mathbf{x}},t}[g(t)^2 \|\mathbf{s}_\theta(\hat{\mathbf{x}}_t, t, \varnothing) - \nabla_{\mathbf{x}'} \log p_t(\mathbf{x}'|\hat{\mathbf{x}})\|_2^2]$ using randomly sampled t and ϵ;

  Update purified $\hat{\mathbf{x}} = \hat{\mathbf{x}} - \mu \frac{\mathrm{g}}{\|\mathrm{g}\|^2}$;

for $c \in \{1, 2, \ldots, K\}$ **do**

  Initial: $\mathbf{x_0} \leftarrow \hat{\mathbf{x}}$

  for $t \in \{0, 1/T, \ldots, 1\}$ **do**

    $\epsilon \sim \mathcal{N}(0, 1)$

    $\mu_t, \sigma_t \leftarrow \mathcal{N}(\mathbf{x_0} * f(\mathbf{x_t}, t), g(t)^2 I)$

    $\mathbf{x}'_t \leftarrow \mu_t + \epsilon * \sigma_t$

    $\nabla_{\mathbf{x}'_t} f(\mathbf{x}'_t, t) \leftarrow \mathbb{E}[\epsilon^T \nabla_{\mathbf{x}'_t} f(\mathbf{x}'_t, t)\epsilon]$

    $\nabla_{\mathbf{x}'_t} \log p(\mathbf{x}'_t|\mathbf{x}_0) \leftarrow -\frac{(\mathbf{x}'_t - \mu_t)}{\sigma_t^2}$

    $\tilde{f}_{DSM}(t) \leftarrow \tilde{f}_{DSM}(t) + g(t)^2 \|\mathbf{s}_\theta(\mathbf{x}'_t, t, \mathbf{c}) + \frac{(\mathbf{x}'_t - \mu_t)}{\sigma_t^2}\|_2^2 - g(t)^2 \|\frac{(\mathbf{x}'_t - \mu_t)}{\sigma_t^2}\|_2^2 - 2\mathbb{E}[\epsilon^T \nabla_{\mathbf{x}'_t} f(\mathbf{x}'_t, t)\epsilon]$

  Approximate $-\log p_\theta(\mathbf{x_{data}}|c) \leftarrow$ Eq.(6)

  Calculate $p_\theta(c|\mathbf{x_{data}}) \leftarrow$ Eq.(8)

return $\tilde{c} \leftarrow \arg\max_c p_\theta(c|\mathbf{x_{data}})$

IV. Experimental results and discussions

A. Experimental dataset acquisition and prepossessing

In our previous work, a hyperspectral dataset of 56 honey products and 21 botanical sources in New Zealand was compiled [7]. Honey products were captured by the SOC710-VP hyperspectral camera, where the spectral wavelength ranges from 400 to 1000 nm. 128 spectral bands were captured in each hyperspectral image, with an approximate 4.9 nm spectral resolution and 520×696 pixels spatial resolution. Subsequently, interested regions of honey were extracted and segmented into 25 (5x5) pieces, with 10x10 pixels in each piece. Within this dataset, clover and manuka honey graded from UMF 5+ to 22+ were conducted to mix with syrups at different levels. The ratios of mixing syrups and honey were intentionally designed into four (5%, 10%, 25%, and 50%). Dynamic calibrations and Savitzky-Golay filters were conducted and averaged on raw hyperspectral data to remove the random variations caused by unstable environmental conditions. Statistically, the dataset was divided into three parts: training, validation, and testing, accounting for 3,640, 1,600, and 1,310 samples, respectively.

B. Spectral classification results on perturbed and pure data

Firstly, we are looking at the results of using CDM to classify the honey spectrum without adulterated shifts. This experiment is analogous to a conventional test, assessing the results of the classification model with consistent conditions between the training and test sets. The test results in the table show that the base classification model is not the most accurate in classifying adulterated honey and pure honey under specific concentrations, 97.7%, and 94.0%. On the contrary, the highest precision and recall scores go to the ANN-based classification, where they are both over 99.0% in Table IV-B. Secondly, the diffusion classifiers slightly drop behind the CNN and RNN-based models with observed adulteration ratios in these two tasks. Surprisingly, a simple LSVM yields the second-best result via all concentration conditions, namely 96.8% precision and 95.1% recall. In actuality, we observe a slight improvement, even a decrease in classification accuracy for the proposed method when testing honey data is perturb-free; the interpretive explanation will be illustrated in Section IV-C. As demonstrated in Table IV-B, a series of generative model-based classifiers are superior to the conventional discriminative model. Numerically, the score-based classifier reaches the highest result, even though it suffers from two different levels of concentration perturbations, from 85.8% to 95.7%. On the contrary, a three-layer ANN struggles to classify those fake samples with concentration perturbations, dropping to nearly 55.3% for precision and 0.0% for recall. Compared with the proposed approach, the VECAE and GANs-based adversarial training shows worse robustness and accuracy against perturbed unseen concentrations, lower than 70.0%. Notably, our proposed method is limited to the tolerance of defending against attacks of adulteration perturbations. Collectively, once the adulterated concentration has shifted over 20% away from the training concentration, our proposed detection model will be penalized for confusing the authentic and adulterated honey. For instance, the accuracy and recall of recognizing the fake honey with concentration 5% and 10% decrease overall by 18.6% and 20.3%, respectively, compared with training via 25% and 50%, but testing with consistent 25% and 50%.

C. Discussion

Results in Table IV-B imply that our proposed method empirically outperforms discriminative methods, particularly when exit shifts on honey adulterated ratios exist. We observe the mean precision and recall increase of 18.3% and 14.0% in the first two columns in Table IV-B. However, we will draw a different conclusion when the table's full columns are assessed collectively. Statistically, a slight decrease is visible in the experimental results when testing a CDM-based classifier with clean data. This phenomenon is probably because the DC-based training process concentrates more on

TABLE I
CLASSIFICATION RESULTS OF DIFFERENT METHODS ON CLEAN DATA AND PERTURBED DATA

Adulterated ratios (train)	Method	Classification results (clean)							
		Adulterated ratios (test)							
		5%		10%		25%		50%	
		prec	recall	prec	recall	prec	recall	prec	recall
25%+50%	LSVM	48.6	0.0	65.7	54.6	99.3	100.0	100.0	100.0
	PLSDA	54.0	43.8	50.0	0.0	99.8	100.0	99.8	96.7
	ANN	55.3	4.3	65.2	52.8	99.9	100.0	100.0	100.0
	CNN	59.3	49.1	68.9	57.0	96.0	98.5	99.1	100.0
	VECAE	65.6	55.7	72.5	64.9	95.5	91.6	96.0	92.6
	GANs	72.5	67.4	77.2	70.3	**99.8**	**100.0**	**99.9**	**100.0**
	CDM	**75.3**	**69.0**	**79.4**	**73.6**	99.3	98.9	97.7	94.0

increasing robustness during classification under unobserved attacks. Unfortunately, more robustness will be gained if the penalty of dropping classification accuracy is paid from a 'no free lunch' theory.

This part will analyze the underlying relationship between classification results and different adulterated concentrations. The lower the concentration values, the worse the results conventional discriminative classifiers will obtain. The reason will be very intuitive since there are slight differences existing in pure samples and adulterated samples with a lower ratio. Slight spectral difference between lower ratios adulterated and pure honey, which poses challenges on our model to make correct decisions.

In this study, the robustness of our conducted spectral model deserves to be more intentionally assessed and analyzed. As mentioned in Section II-C, the Lipschitiz constant assesses the smoothing ability against the input perturbations. The lower the value, the more robust the model will defend the perturbations from spectral inputs [17]. Accordingly, the certified robustness of the base classifier can be analyzed from the perspective of randomized smoothing (Chen et al. [19]). To analyze the Lipschitz constant of our classification models, we must dive them according to the chain rule since the ELBO function $g(.)$ of CDM is wrapped up with the Bayesian reverse. Thus, the first term is a derivation of the softmax function.

$$
\begin{aligned}
&\max_{\mathbf{u}} \mathbf{u}^\top \frac{\partial}{\partial \mathbf{x}} p(y|\mathbf{x}) \\
&= \| \frac{\partial}{\partial \mathbf{x}} p(y|\mathbf{x}) \|_2 \\
&= \| \frac{\partial}{\partial \mathbf{x}} \mathrm{softmax}(g(\mathbf{x})) \|_2 \\
&= \| \frac{\partial}{\partial \mathbf{x}} \mathrm{softmax}(\mathbb{E}_\epsilon \left[f(\mathbf{x} + \sigma_t \epsilon) \right]) \|_2 \\
&\leq \sum_{y=1}^{K} \| \frac{\partial \mathrm{softmax}(\mathbb{E}_\epsilon [f(\mathbf{x} + \sigma_t \epsilon)])}{\partial (-\frac{1}{DT} \mathbb{E}_\epsilon \left[f(\mathbf{x} + \sigma_t \epsilon) \right])} \|_2 \| \frac{\partial (-\frac{1}{DT} \mathbb{E}_\epsilon \left[f(\mathbf{x} + \sigma_t \epsilon) \right])}{\partial \mathbf{x}} \| \\
&\leq \frac{1}{2\sqrt{2}} \| \frac{\partial (-\frac{1}{DT} \mathbb{E}_\epsilon \left[f(\mathbf{x} + \epsilon) \right])}{\partial \mathbf{x}} \|_2
\end{aligned}
\tag{9}
$$

As for the second term, we can make an analogous for the upper bound of CDM in Eq.6 as a smoothing function to follow the format of Eq.2.

$$
\begin{aligned}
&\| \frac{\partial (\mathbb{E}_\epsilon \left[f(\mathbf{x} + \epsilon) \right])}{\partial \mathbf{x}} \|_2 \\
&= \| \frac{\partial (\lambda(t) \mathbb{E}_{\mathbf{x},t} [\| \mathbf{s}_\theta(\mathbf{x}'_t, t) - \nabla_{\mathbf{x}'} \log p_t(\mathbf{x}'|\mathbf{x}) \|_2^2])}{\partial (\mathbf{x})} \|_2 \\
&\leq \frac{1}{T} \sum_{t=1}^{T} \frac{\lambda(t)}{g(t)} (D \sqrt{\frac{2}{\pi}} + 2\sqrt{D})
\end{aligned}
\tag{10}
$$

On top of this, the Lipschitz constant if the loss term $\| \mathbf{s}_\theta(\mathbf{x}_t, t) - \nabla_{\mathbf{x}'} \log p_t(\mathbf{x}'|\mathbf{x}) \|_2^2$ and $g(t)^2$ are normalized as $[0, 1]^D$, which is nearly identical to that in the "weak law" of randomized smoothing.

$$
\| f_\theta(\mathbf{x} + \boldsymbol{\epsilon})_y - f_\theta(\mathbf{x})_y \| \leq \frac{1}{2\sqrt{2}} \sum_{t=1}^{T} \frac{\lambda(t)}{g(t) T} (\sqrt{\frac{2}{\pi}} + \frac{2}{\sqrt{D}}) \| \boldsymbol{\epsilon} \|
\tag{11}
$$

Accordingly, the specific value of the derived Lipschitz constant K_0 has been tested in the work of Chen et al. [19], suggesting that the CDM-based classifier is smaller than 0.02 on the CIFAR-10 dataset.

Although our study does not further implement the most advanced techniques in generative models, like EDM [20], using generative models to accelerate the defensive robustness of the AI-based spectral detection method against food adulteration has proven feasible. Moreover, our proposed method has a defensive upper bound against perturbation attacks, and the model tends to collapse when the mixed syrup-honey ratio shifts over 20%. However, this upper bound comes from the empirical experiments; thereby, a more accurate upper bound needs to be determined by testing the models' defensive performance against attacks from different scales of adulterated concentration values. The aforementioned method will be used to update our results in future works. Moreover, our experiment using honey origins from New Zealand is very limited, complying with the New Zealand food policy. In the future, we will collect honey from more geographical sources to verify our results.

V. CONCLUSION

In this study, we propose a new strategy to classify honey authenticity using hyperspectral data robustly. Conditional diffusion models are further wrapped with the Bayesian rule

to handle those perturbations from shifted adulterants. Theoretically, the probabilistic classifier conducted in our study is certified robust according to the analysis of the Lipschitz constant against input noise. Empirically, our experimental results imply that our proposed method effectively reduces the uncertainty due to adulterated perturbations, enabling our spectral model to classify adulterated honey even if the adulterants are shifted. In the future, we hope our findings can contribute to food fraud detection using spectroscopic techniques, improving the robustness of detection models when challenged by adulteration perturbations designed by dishonest producers.

REFERENCES

[1] T. Phillips, W. Abdulla, A new honey adulteration detection approach using hyperspectral imaging and machine learning, European Food Res. Technol. 249 (2023) 259–272. doi:10.1007/s00217-022-04113-9.

[2] G. Zhang, W. Abdulla, On honey authentication and adulterant detection techniques, Food Control 138 (2022) 108992. doi:10.1016/j.foodcont.2022.108992.

[3] A. Naila, S. H. Flint, A. Sulaiman, A. Ajit, Z. Weeds, Classical and novel approaches to the analysis of honey and detection of adulterants, Food Control 90 (2018) 152–165. doi:10.1016/j.foodcont.2018.02.027.

[4] W. Skaff, R. El Hajj, L. Hanna-Wakim, N. Estephan, Detection of adulteration in honey by infrared spectroscopy and chemometrics: Effect on human health, J. Food Process. and Preserv. 46 (10) (2022) e15438. doi:10.1111/jfpp.15438.

[5] M. Ferreiro-González, E. Espada-Bellido, L. Guillén-Cueto, M. Palma, C. G. Barroso, G. F. Barbero, Rapid quantification of honey adulteration by visible-near infrared spectroscopy combined with chemometrics, Talanta 188 (2018) 288–292. doi:10.1016/j.talanta.2018.05.095.

[6] J. Cheng, G. Zhang, W. Abdulla, J. Sun, Advancing fraud detection in new zealand mānuka honey: Integrating hyperspectral imaging and ganomaly-based one-class classification, Food Biosci. 60 (2024) 104428. doi:10.1016/j.fbio.2024.104428.

[7] T. Phillips, W. Abdulla, Variational autoencoders for generating hyperspectral imaging honey adulteration data, in: 2022 IEEE/CVF Conf. on Comp. Vis. and Pattern Recognit. Workshops (CVPRW), 2022, pp. 213–220. doi:10.1109/CVPRW56347.2022.00035.

[8] E. Ahmed, Detection of honey adulteration using machine learning, PLOS Digital Health 3 (6) (2024) 1–25. doi:10.1371/journal.pdig.0000536.

[9] B. Yang, C. Chen, F. Chen, C. Chen, J. Tang, R. Gao, X. Lv, Identification of cumin and fennel from different regions based on generative adversarial networks and near infrared spectroscopy, Spectr. Acta Part A: Mol. and Biomol. Spectrosc. 260 (2021) 119956. doi:10.1016/j.saa.2021.119956.

[10] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, B. Poole, Score-based generative modeling through stochastic differential equations, in: Int. Conf. on Learn. Rep. (ICLR), 2021.

[11] Y. Song, C. Durkan, I. Murray, S. Ermon, Maximum likelihood training of score-based diffusion models, in: A. Beygelzimer, Y. Dauphin, P. Liang, J. W. Vaughan (Eds.), Advances in Neural Information Processing Systems, 2021.

[12] J. Ho, T. Salimans, Classifier-free diffusion guidance, in: NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications, 2021.

[13] X. Wu, B. Xu, H. Luo, R. Ma, Z. Du, X. Zhang, H. Liu, Y. Zhang, Adulteration quantification of cheap honey in high-quality manuka honey by two-dimensional correlation spectroscopy combined with deep learning, Food Control 154 (2023) 110010. doi:10.1016/j.foodcont.2023.110010.

[14] L. Bruce, C. Koger, J. Li, Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction, IEEE Trans. on Geosci. Remote Sens. 40 (10) (2002) 2331–2338. doi:10.1109/TGRS.2002.804721.

[15] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, et al., Deep learning applications and challenges in big data analytics, Journal of Big Data 2 (1) (2015) 1. doi:10.1186/s40537-014-0007-7.

[16] L. Yang, Z. Zhang, Y. Song, S. Hong, R. Xu, Y. Zhao, W. Zhang, B. Cui, M.-H. Yang, Diffusion models: A comprehensive survey of methods and applications, ACM Comput. Surv. 56 (4) (Nov. 2023). doi:10.1145/3626235.

[17] A. Virmaux, K. Scaman, Lipschitz regularity of deep neural networks: analysis and efficient estimation, in: S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett (Eds.), Advances in Neural Information Processing Systems, Vol. 31, Curran Associates, Inc., 2018.

[18] A. M. Rekavandi, F. Farokhi, O. Ohrimenko, B. I. P. Rubinstein, Certified adversarial robustness via randomized α-smoothing for regression models, in: The Thirty-eighth Annual Conference on Neural Information Processing Systems, 2024.

[19] H. Chen, Y. Dong, Z. Wang, X. Yang, C. Duan, H. Su, J. Zhu, Robust classification via a single diffusion model, in: Forty-first International Conference on Machine Learning, 2024.

[20] T. Karras, M. Aittala, T. Aila, S. Laine, Elucidating the design space of diffusion-based generative models, in: A. H. Oh, A. Agarwal, D. Belgrave, K. Cho (Eds.), Advances in Neural Information Processing Systems, 2022.