THESIS

NONLINEAR DYNAMICS AND MACHINE LEARNING CLASSIFICATION OF PLANT PIGMENT PATTERNS

Submitted by
Kaylee Wong Dolloff
Department of Mathematics

In partial fulfillment of the requirements
For the Degree of Master of Science
Colorado State University
Fort Collins, Colorado
Summer 2023

Master's Committee:

Advisor: Patrick Shipman

Jennifer Mueller
Joe von Fischer

ABSTRACT

NONLINEAR DYNAMICS AND MACHINE LEARNING CLASSIFICATION OF PLANT PIGMENT PATTERNS

Plants exhibit a variety of vibrant colors that are both beautiful and functional. They owe their reds, purples, and blues to a class of pigments called anthocyanins. Many plants possess spatial variation in their anthocyanin concentration and color, which manifest as diverse patterns on their leaves and flowers. Flower patterns can influence interactions with pollinators, who may have innate preferences for certain patterns and can learn to distinguish between them. Recent work has identified the genes and proteins involved in activation and inhibition of anthocyanin synthesis in some species of *Mimulus* and showed that their dynamics can be described with a two-component diffusion model. In this thesis, we combine numerical simulations of this model with machine-learning algorithms to classify patterns based on a parameter value that influences the pattern spot size and density. A key challenge is to successfully classify using 2-dimensional spot data, which would permit the classification of real petal data from photos. Our approach makes use of the Voronoi mountain function to construct a 3-dimensional surface from the 2-dimensional data. Classification is very successful with simulated data, and it produces plausible results for real *Mimulus* petals.

ACKNOWLEDGEMENTS

First, I would like to express my immense, uncountably infinite gratitude to my advisor, Patrick Shipman. This thesis wouldn't have been possible without his time, patience, and encouragement. His optimism and good humor have made my time in graduate school incredibly enjoyable, and I count myself very fortunate to have gotten to work with him.

I would also like to thank my committee members, Jennifer Mueller and Joe von Fischer. I learned so much in their classes and had a wonderful time doing it. I'm grateful for their academic and moral support, as well as their input on this thesis.

My experience at CSU would not have been the same without the company of such a terrific math department and graduate students. In particular, I would like to express my appreciation to everyone in Weber 234 for their friendship and support. I couldn't ask for better people to share an office with. Thank you as well to my friends and mentors in the math department at UCCS, who encouraged me to study math and pursue a graduate degree to begin with.

Last and certainly not least, a massive thank you to my family for unconditionally loving, supporting, and believing in me (and, in the case of my parents, driving 3 hours to Fort Collins to visit me every month). Their encouragement helped me through difficult days, and I wouldn't be here without them.

DEDICATION

I would like to dedicate this thesis to my family.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Plants are a diverse and vibrant group of organisms, owing their colors to a variety of cell pigments. The most well-known of these pigments is chlorophyll, which gives plants their green color. The other groups of pigments are carotenoids (yellow, orange, and red), betalains (red, purple, yellow, and orange), and anthocyanins (red, purple, and blue) [1]. Plant pigments serve a variety of functions, including photoprotection and cell defense [1–3]. They also influence and are influenced by cell pH, a vital regulator of cell processes [4, 5].

Anthocyanins are common in many plants, though often not in quantities large enough to produce color. They are responsible for the reds and purples of autumn foliage and are often a dominant presence in young leaves, which gradually turn green as they mature [6]. Some plants have colored leaves year-round, which can often be attributed to high concentrations of anthocyanins. Many types of flowers owe their colors to the presence of anthocyanins, as do some red and purple fruits and vegetables [7–9].

Figure 1.1: Samples from the vibrant palette produced by anthocyanins. Top row: An autumn leaf, red onions (*Allium cepa*), blackberries (*Rubus* sp.). Bottom row: Dahlia flowers (*Dahlia* hybrid), petunia flowers (*Petunia* × *hybrida*), young begonia leaf (*Begonia rex*).

Anthocyanins are of particular interest to the food and beverage industry due to their common presence in edible plants and their byproducts. The colors imparted by anthocyanins influence consumer choices of produce and beverages [9–11]; one such bevarage is wine, whose color and taste come from the contributions of anthocyanins in grapes [12]. Anthocyanins also have the potential to be used as natural colorants, which requires an understanding of their behavior and stability outside of the cell [13, 14].

Inside the cell, anthocyanins display an array of unique behaviors. Microscopy of floral epidermal cells (Figure 1.2) shows that anthocyanins associate with other molecules to form complexes (panel a) or themselves to form particles (b,c). Anthocyanins in solution form particles as the solution evaporates (d); the particles fluoresce when exposed to ultraviolet light (e).

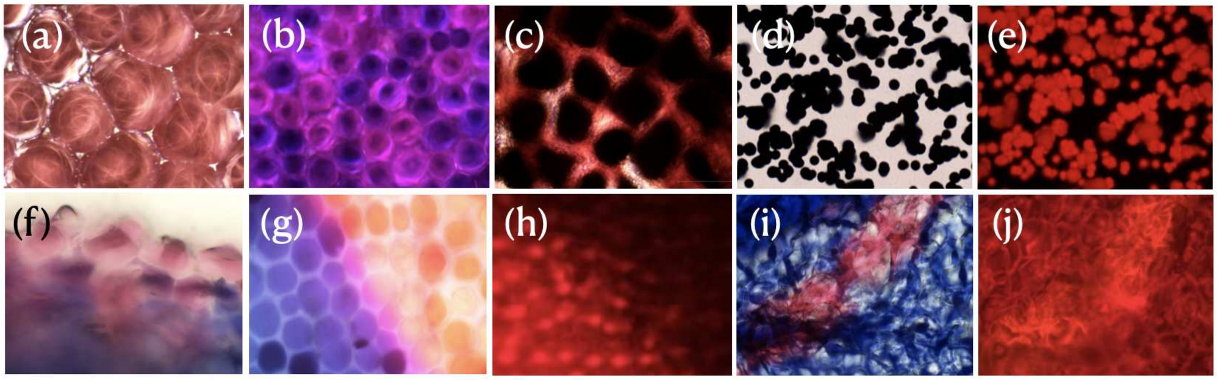


Figure 1.2: Microscope images of pigmented plant cells and anthocyanin complexes. (a) Anthocyanins associated with microtubules in geranium cells. (b,c) Various phases of anthocyanin complexes are present in these cells of (b) purple petunia and (c) blueberry. (d) These black particles remained after an acidic solution of the anthocyanin cyanidin 3-glucoside evaporated. (e) The particles of panel (d) observed in fluorescent light. (f) Conical, blue cells of a butterfly pea flower turn red upon being exposed to (acidic) HCl; (g) HCl diffused into a live flower epidermal peel turned the originally blue cells orange; in this image, NH3 (a base) is diffusing over the peel from the left, turning the cells first pink and then back to blue. (h) The cells of panel (g) exposed to fluorescent light; the blue cells fluoresce; the orange ones do not. (i) The vein cells of a blue butterfly pea flower turn red upon exposure to HCl. (j) A fluorescence image of the cells in panel (i) reveals that the red cells fluoresce. Samples were prepared and imaged by S. Thompson.

Interestingly, when the epidermal cells in Figure 1.2 are treated with acids and bases, color changes and fluorescence occur (panels f-j). The dramatic pH-induced color changes seen in anthocyanins are not observed in any other biological pigment. Figure 1.3 shows the color variation of cabbage and radish extracts, which contain a variety of anthocyanins, and pure malvin, the predominant anthocyanin in red wine.

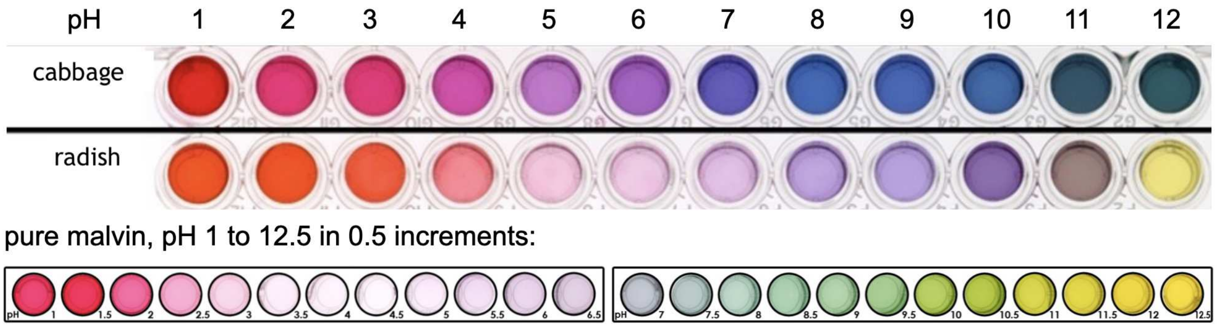


Figure 1.3: Anthocyanin extracts from cabbage and radish, and pure malvin, at a range of pH values. Extracts prepared by S. Thompson [15].

As we see in Figure 1.1, anthocyanins contribute to the diverse colors seen on many flowers. Differing anthocyanin concentrations between petal cells can also result in pigmented patterns. Flower patterns are diverse, and they can even vary among members of the same species; ornamental plant growers have taken advantage of this variation to produce a wide variety of flower cultivars. Many species of monkeyflower, genus *Mimulus*, have particularly distinctive anthocyanin spot patterns. These spots affect pollinator visitation and may be the target of pollinator-mediated selection [16]. Natural variation in spotting in *Mimulus luteus* is shown in Figure 1.4.

Figure 1.4: Pattern variation in *Mimulus luteus* flowers. Photos from [16].

Baoqing Ding and colleagues found that the activator-inhibitor pair that controls anthocyanin spot production in *Mimulus lewisii* and *Mimulus guttatus* can be modelled using a system of 2 partial differential equations known as the Gierer-Meinhardt reaction diffusion model [17]. The

Gierer-Meinhardt model requires an activator and an inhibitor, which diffuse to nearby cells at different rates. Additionally, the activator must activate both itself and its inhibitor, and the inhibitor must travel faster than the activator. If both of these conditions are met, the Gierer-Meinhardt model guarantees pattern formation. Ding and colleagues verified the model for *M. lewisii* pattern formation by using it to generate simulated patterns. The resulting simulations resembled the spots found on real petals (Figure 1.5).

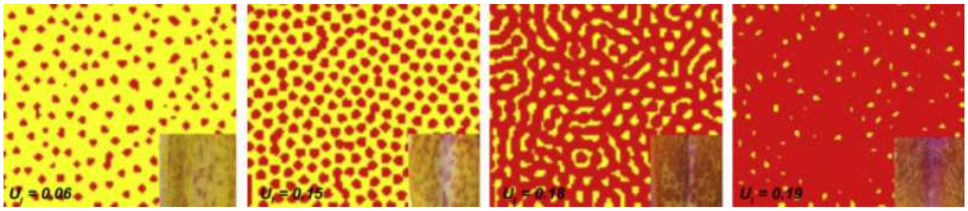


Figure 1.5: Pattern variation in genetically modified *Mimulus lewisii* flowers, compared with simulated patterns. Figure from [17].

In this thesis, we will build on the work of Ding and colleagues by combining the Gierer-Meinhardt model with machine learning to develop a method for anthocyanin pigment pattern classification that can be used with real flower petals. In chapter 2, we will first provide background information on the structure and function of anthocyanins in flowers, then introduce the Gierer-Meinhardt model for anthocyanin spot patterns in *M. lewisii*. In chapter 3, we will present our methods and results for pattern classification.

Chapter 2

Anthocyanins and flower patterns

2.1 Anthocyanin structure

The chemical structure of anthocyanins is integral to their color-changing abilities. Anthocyanins are formed from a parent molecule, the flavylium cation (Figure 2.1, column (a)). The rings, labelled A, B, and C in the figure, form the chromophore. The addition of $-$H, $-$OH (hydroxy), or $-O-CH_3$ (methoxy) groups to the positions labelled R_i transforms the flavylium cation into one of the 20 known types of anthocyanidins. The anthocyanidins pelargonidin and malvidin are shown in column (b) of Figure 2.1; note the differences in functional groups in the R_i positions.

Anthocyanidins are water-insoluble. The addition of sugars (or glycosyl units, shown in column (c) of Figure 2.1) at some of the positions marked 1-8 or 1’-6’ turns an anthocyanidin into one of the over 900 known types of anthocyanins, which are water-soluble. Adding glycosyl units to positions 3 and 5 on malvidin produces the anthocyanin malvin. Anthocyanins can be modified further through the addition of acyl units to the glycosyl units, shown in column (d). Acyl units can donate electrons to the chromophore, which changes the geometry and color of the molecule.

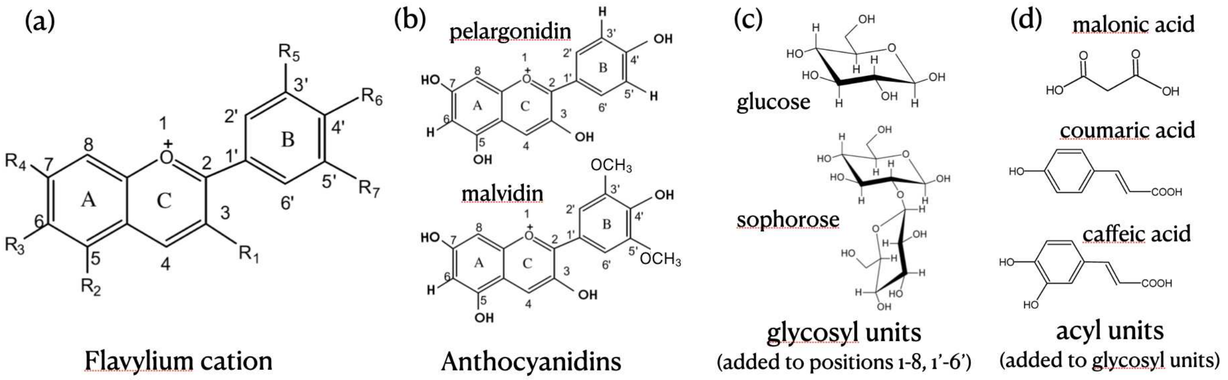


Figure 2.1: Column (a): the flavylium cation; column (b): examples of anthocyanidins produced by substitutions at the positions R_i; column (c): glycosyl (sugar) units such as glucose and sophorose are added to positions 1-8 and 1’-6’ to produce a water-soluble anthocyanin; column (d): acyl units may be added to glycosyl units. Figure from [15].

An anthocyanin can undergo a series of fast and slow reactions to form different chemical species (Figure 2.2) [4]: A$^-$, the anhydrobase anion, appears blue. The A$^-$ species undergoes a fast acid-base reaction with a proton that transforms it into the purple quinoidal base, A, and the addition of another proton transforms A into the red flavylium cation, AH$^+$. The slow hydration of AH$^+$ transforms it into the colorless hemiketal, B. B is a structural isomer (or tautomer) of the light yellow *cis*-chalcone Cc, meaning it has the same number of atoms of each element, but with a distinct combination of bonds. Cc transforms into the *trans*-chalcone Ct through slow isomerization, where the functional groups of the molecule shift their spatial arrangement.

The equilibria, and thus the dominant species, of these reactions are influenced by the pH of the solution, which then determines the color of the solution. In particular, higher pH values favor reactions to the left of Figure 2.2. In neutral and alkaline solutions (pH ≥ 7), the blue A$^-$ species dominates. In solutions with $4 <$ pH ≤ 6, which is the typical vacuolar pH range, the purple A and red AH$^+$ species dominate. The colorless B species dominates in solutions with pH 4.

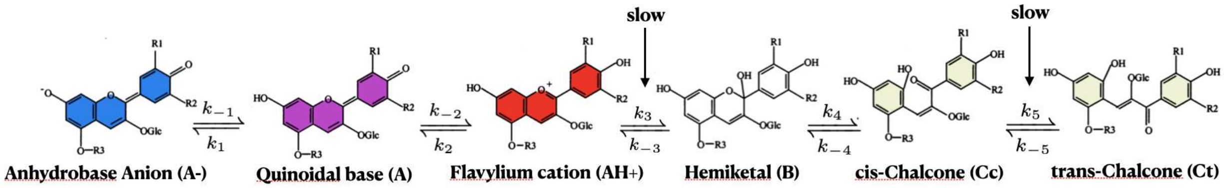


Figure 2.2: pH-dependent anthocyanin structural changes. Figure from [15]

Single anthocyanin molecules, referred to here as monomers, can associate with each other and other molecules [4, 5, 18–22]. Intramolecular anthocyanin associations are called dimers, trimers, and j-mers (for $j \geq 3$) and are facilitated through multiple types of chemical bonding. The type of bond, its strength, and its geometry are factors in anthocyanin association [4]; for example, AH$^+$ has a positive charge, so it can ion-pair with negatively-charged ions to create zwitterionic molecules (AH$^+$)$^-$. Similarly, A$^-$ can pair with positively-charged ions. A is less soluble because it has a neutral charge, which results in a reduction in association ability.

2.2 Anthocyanin reaction scheme

The basic anthocyanin species reaction scheme is as follows:

$$\begin{array}{llllll}
A^- + H^+ \underset{k_1}{\overset{k_{-1}}{\rightleftharpoons}} & A & & & & \\
 & A & +AH^+, & B & \underset{k_{-6}}{\overset{k_6}{\rightleftharpoons}} & B^- + H^+ \\
 & & AH^+ \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} & B & & +H^+ \\
 & & & B & \underset{k_{-4}}{\overset{k_4}{\rightleftharpoons}} C \underset{k_{-5}}{\overset{k_5}{\rightleftharpoons}} C^- & +H^+
\end{array} \tag{2.1}$$

We can use this scheme and the law of mass action to write a system of differential equations for the change in concentration of each species:

$$\frac{d}{dt}[A^-] = \{-k_{-1}[A^-][H^+] + k_1[A]\} \tag{2.2}$$

$$\frac{d}{dt}[A] = -\{-k_{-1}[A^-][H^+] + k_1[A]\} + \{-k_{-2}[A][H^+] + k_2[AH^+]\} \tag{2.3}$$

$$\frac{d}{dt}[AH^+] = -\{-k_{-2}[A][H^+] + k_2[AH^+]\} + \{-k_3[AH^+] + k_{-3}[B][H^+]\} \tag{2.4}$$

$$\frac{d}{dt}[B] = -\{-k_3[AH^+] + k_{-3}[B][H^+]\} + \{-k_4[B] + k_{-4}[C]\} + \{-k_6[B] + k_{-6}[B^-][H^+]\} \tag{2.5}$$

$$\frac{d}{dt}[B^-] = -\{-k_6[B] + k_{-6}[B^-][H^+]\} \tag{2.6}$$

$$\frac{d}{dt}[C] = -\{-k_4[B] + k_{-4}[C]\} + \{-k_5[C] + k_{-5}[C^-][H^+]\} \tag{2.7}$$

$$\frac{d}{dt}[C^-] = -\{-k_5[C] + k_{-5}[C^-][H^+]\} \tag{2.8}$$

The equilibrium constants are the ratios of the forward and backward reaction rate constants. Equilibrium constants are known for each type of anthocyanin.

$$K_1 = \frac{k_1}{k_{-1}}, K_2 = \frac{k_2}{k_{-2}}, K_3 = \frac{k_3}{k_{-3}}, K_4 = \frac{k_4}{k_{-4}}, K_5 = \frac{k_5}{k_{-5}}, K_6 = \frac{k_6}{k_{-6}}. \tag{2.9}$$

Equilibrium solutions can be found by setting the above time derivatives equal to 0 and writing each concentration in terms of [A].

$$\begin{aligned}
[A^-] &= \frac{K_1}{[H^+]}[A] \\
[A] &= [A] \\
[AH^+] &= \frac{[H^+]}{K_2}[A] \\
[B] &= \frac{K_3}{K_2}[A] \\
[B^-] &= \frac{K_6 K_3}{[H^+]K_2}[A] \\
[C] &= \frac{K_4 K_3}{K2}[A] \\
[C^-] &= \frac{K_5 K_4 K_3}{K_2[H^+]}[A]
\end{aligned} \tag{2.10}$$

The total anthocyanin concentration $[T]$ is the sum of the concentration of each species and is a conserved quantity.

$$[T] = [A^-] + [A] + [AH^+] + [B] + [B^-] + [C] + [C^-] \tag{2.11}$$

Substituting in the quantities from (2.10) gives

$$[T] = K_T([H^+])[A] \tag{2.12}$$

where $K_T([H^+])$ is a function of the concentration of hydrogen ions:

$$K_T([H^+]) = \frac{K_1}{[H^+]} + 1 + \frac{[H^+]}{K_2} + \frac{K_3}{K_2} + \frac{K_6 K_3}{[H^+]K_2} + \frac{K_4 K_3}{K2} + \frac{K_5 K_4 K_3}{K_2[H^+]} \tag{2.13}$$

A graph of anthocyanin steady-state mole fractions as a function of pH is pictured in Figure 2.3. Anthocyanins are stored in the vacuole. Typical vacuolar pH ranges from 4 to 6, so the equilibrium solution of the basic anthocyanin reaction scheme indicates that the colorless B species should dominate. In spite of this conclusion, colored anthocyanin species are still observed in plant cells (Figure 1.2).

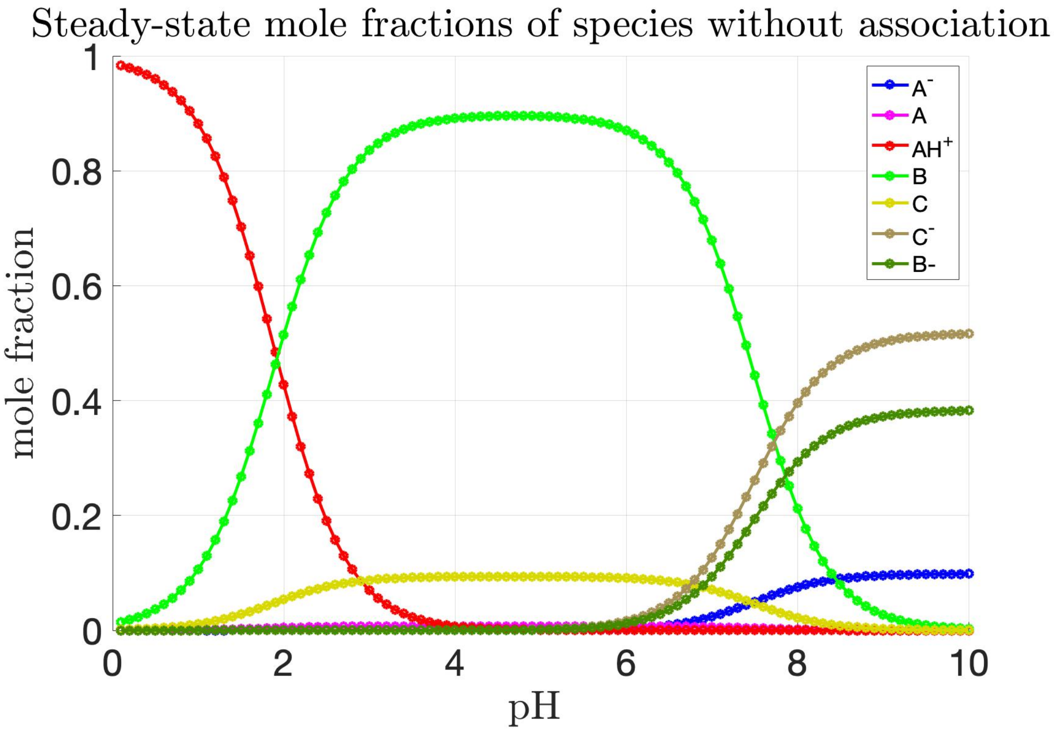


Figure 2.3: Equilibrium mole fractions for each anthocyanin species as a function of pH.

One hypothesis is that association can cause chemical changes that allow anthocyanins to display their colors in the cell. For example, anthocyanins formed from the anthocyanidins cyanidin, petunidin, and delphinidin have two -OH groups on their B ring that can form both metal and non-metal complexes, which exhibit unusual free-radical and excited-state behavior. These properties help explain the vibrant blue color seen in flowers like the hydrangea, whose anthocyanins form complexes with aluminum ions in acidic soil, and the blue cornflower, *Centarurea cyanus*, whose anthocyanins form complexes with iron, magnesium, and calcium ions [14, 23–25]. Hydrangeas are unique because they have a wide range of pink to blue coloration that is entirely derived from one anthocyanin, delphinidin 3-glucoside (Figure 2.4) [26].

Figure 2.4: Two *Hydrangea macrophylla* cultivars with different petal colors, both a product of delphinidin 3-glucoside.

2.3 Flower color

Pollinators have the ability to distinguish between flowers based on color and other morphological properties, and pollinator-mediated selection on flowers can be strong [27]. The vision of bees has been studied extensively; the honeybee was even the first non-human animal that was shown to have color vision in experiments conducted by Karl von Frisch [28, 29]. Von Frisch demonstrated that honeybees use the color, rather than brightness, of stimuli to discriminate between rewards.

Like humans and other primates, bees have trichromatic vision. Figure 2.5 shows examples of flowers viewed by humans and bees. Bees can detect ultraviolet (UV) [30], and their photoreceptor sensitivities peak in the UV, blue, and green areas of the spectrum [31–33]. The three photoreceptor types are S, short-wavelength, peaking at approximately 344 nm (UV), M, middle-wavelength, peaking at approximately 436 nm (blue), and L, long-wavelength, peaking at approximately 544 nm (green) [34]. In human cone photoreceptors, S peaks at approximately 420 nm (blue), M peaks at approximately 534 nm (green), and L peaks at approximately 564 nm (red). The rods peak at approximately 498 nm [35]. The similarities of color vision in primates and bees appear to be a result of convergent evolution [36].

Bees also possess a mechanism for color constancy, the ability to identify colors in the same hue regardless of lighting conditions [36, 37]. Without color constancy, bees would lack the ability

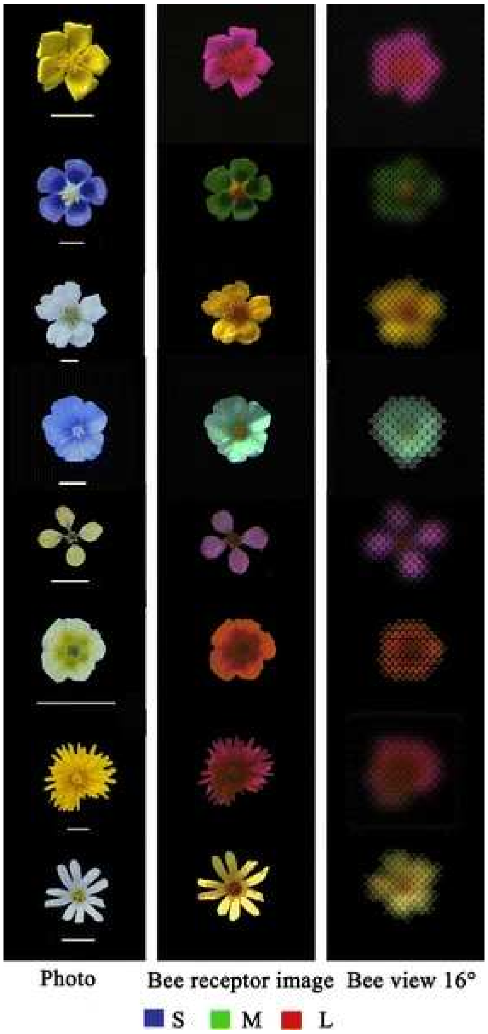


Figure 2.5: From left to right, flowers viewed in human vision, bee colors, and projected into the bee vision lattice. Blue corresponds to the S (UV) receptor, green to the M (blue) receptor, and red to the L (green) receptor. Flowers are, top to bottom, *Helianthemum nummularia, Aquilegia vulgaris, Mespilus germanica, Linum austriacum, Vella spinosa, Nonea lutea, Taraxacum officinale, Stellaria holostea*. Figure from [29]

to recognize the same object repeatedly in different conditions, which would prevent them from using color recognition and learning as a component in their foraging decisions.

Indeed, bees can learn to use color in conjunction with other visual, olfactory, or tactile properties to forage efficiently and maximize nectar rewards. Von Frisch's color vision experiments trained honeybees to associate a reward with a card of a specific color [28], and further work has advanced our understanding of what bees are capable of learning. Bees learn very quickly, though the learning speed and success rate varies between colors. In one set of experiments, bees were trained to visit violet (413 and 428 nm) after just one trial, while blue-green (494 nm) was the most difficult color for the bees to learn [29, 38]. Later experiments showed that bees possess very complex and flexible learning mechanisms, contrary to the assumption that insects and other invertebrates have limited (if any) learning capacity. Remarkably, they have different memory-forming phases that allow them to process and use new information before the old information is stored [38].

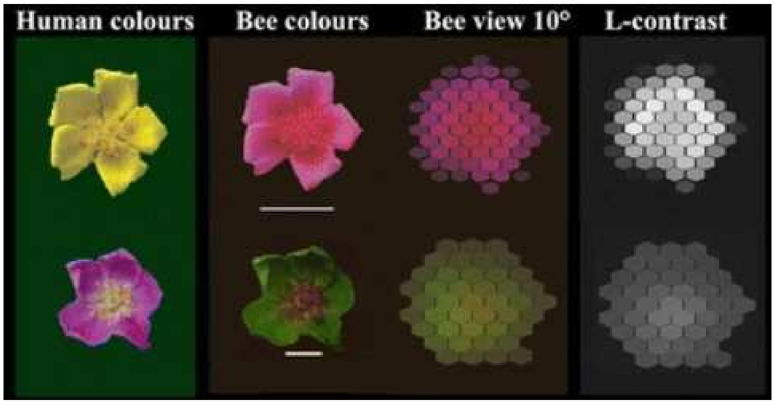
Color is an important factor in the bee learning process, but there is little evidence to suggest that bees have innate color preferences [27]. Furthermore, bees are virtually colorblind when they view flowers at a distance [39,40]; instead, they use their green (L) receptor to perceive brightness, which means the L receptor contrast of a flower compared to its background is important for detection [39, 41]. Bees also have relatively low-resolution distance vision [29], so small flowers with colors offering low L-contrast are difficult for them to detect at a distance. Despite this, many such flowers are bee-pollinated.

2.4 Flower patterns

To make sense of this apparent discrepancy, we must consider the role of flower patterns in pollinator signaling. Many flowers have multicolored patterns on their petals, caused by variations in the concentration of anthocyanins and other pigments. There are a wide variety of patterns that anthocyanins can form, including spots, veins, and complex arrangements of lines (Figure 2.6, panel (a)). Many ornamental flowering plants are selectively bred for distinctive patterning.

Bees are able to discriminate between petal patterns, use prior knowledge of patterns to recognize their variations [42], and, after extensive training, can learn to generalize them [43]. While bees possess advanced pattern discrimination abilities at close range, they are incapable of discerning floral details at a distance due to their poor visual resolution (Figure 2.6, panel (b)). As a result, many experiments that test bee pattern preference at only one distance seem to contradict other experiments that test at a different distance [44]. Different spatial scales must be considered to understand the effect of flower patterns on bees and vice versa.

(a) Anthocyanin patterns on petals: From left to right, patterns on a desert willow, a petunia, a snapdragon vine, a palo colorado, and a moth orchid.



(b) Flowers as viewed by a human and a honeybee, including their L-receptor contrast. Blue corresponds to the S (UV) receptor, green to the M (blue) receptor, and red to the L (green) receptor. Figure from [41]

Figure 2.6: Examples of flower patterns. Panel (a) shows various examples of anthocyanin petal patterns, while panel (b) shows flower patterns as seen by humans and bees.

The low resolution of insect eyes prevents bees from resolving elaborate flower patterns from a distance. Their ability to see patterns is also influenced by their flight trajectory [45]. However, flower patterns can still send long-distance signals even if bees are too far away to fully discern their details. Bees are reliant on L (green) receptor contrast for long-distance pattern detection, so contrasting colors within a flower can greatly enhance its visibility [41]. The presence of spot and ring markings reduce naive bumblebee search time compared to unpatterned flowers [46]. Detection of colored rings is enhanced further depending on the contrast between the inside and outside of the ring. Furthermore, bees detect large single-colored disks better than colored ring markings [47], but many flowers have complex patterns that would seemingly be difficult for a bee to detect from afar.

However, we also need to consider bee vision at a smaller scale. At close range, bees possess complex pattern processing abilities. They can be trained to recognize patterns on rewarding flowers and avoid patterns associated with punishment [46]. Patterns can be attractive to naive bees as well. One category of patterns called nectar guides visually signals the location of nectar in a flower, improving foraging efficiency and ensuring that bees and other pollinators come into contact with pollen. Bees prefer flowers with radiating nectar guide-like patterns to both plain flowers and those with randomly-arranged patterns [44], and the presence of a nectar guide increased foraging speed in both naive and experienced bees [48]. Additionally, moving the nectar guide away from the location of the nectary disrupts bee search times, indicating that the pattern itself and not its position plays a role in guiding bees [49]. Therefore, flower patterns must balance long-range detectability with short-range attractiveness.

Although bee vision has been studied at both long-range and close-range, the process through which bees combine visual information from different scales has yet to be understood. The precise functional significance of flower patterns is also an active area of study.

2.5 Pattern model

The mechanics of pattern formation have long fascinated both mathematicians and biologists, and many models for these mechanisms have been proposed. One of these is the reaction-diffusion model, proposed by Alan Turing [50] and independently advanced by Alfred Gierer and Hans Meinhardt [51, 52]. Turing showed that pattern formation can occur through the interaction of two substances with different diffusion rates (the speed at which the substance spreads). The term "reaction-diffusion" refers to any system with this property, but not all reaction-diffusion systems form patterns. Gierer and Meinhardt's work introduced the conditions that guarantee pattern formation in a reaction-diffusion system [53].

The Gierer-Meinhardt reaction-diffusion model requires two components, an activator and inhibitor. The activator must activate both itself and its inhibitor, the inhibitor must inhibit the activator, and the inhibitor must have a larger diffusion coefficient than the activator [51]. The activator is referred to as "short-range," while the inhibitor is "long-range." Pattern formation occurs in a reaction-diffusion system if and only if the conditions of local activator self-activation and long-range inhibition are satisfied. These assumptions were not included in Turing's original model [53].

The mathematical formulation is a system of partial differential equations:

$$\begin{aligned} \frac{\partial A}{\partial t} &= D_A \Delta A + G_A \left(\frac{A^2}{I} + \widetilde{A_0} \right) - U_A A \\ \frac{\partial I}{\partial t} &= D_I \Delta I + G_I A^2 - U_I I. \end{aligned} \tag{2.14}$$

Here, A is the concentration of activator and I is the concentration of inhibitor. The operator Δ is the 2-dimensional Laplacian, $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$. The constants D_A and D_I are the diffusion coefficients of the activator and inhibitor, respectively. We require $D_A << D_I$ due to the assumption of short-range activation and long-range inhibition. The activator activates itself with potency G_A and the inhibitor with potency G_I. The terms $G_A \widetilde{A_0}$ and $I_0 = 0$ are the background production rates of the activator and inhibitor. The term $G_A \frac{A^2}{I}$ refers to the activation of the activator by itself as well as its inhibition by the inhibitor, and $G_I A^2$ refers to the activation of the inhibitor by the activator. The constants U_A and U_I are the degradation rate constants of the activator and inhibitor. Figure 2.7 shows patterns produced by the Gierer-Meinhardt model with different parameters.

The monkeyflowers, genus *Mimulus*, are used as a model organism for several purposes, including pollinator ecology, speciation, genetics, and evolutionary developmental biology (evo-devo). Many species have distinctive anthocyanin spot patterns. Baoqing Ding and colleagues showed that the activator and inhibitor of *Mimulus lewisii* and *Mimulus guttatus* spot patterns are in the R2R3-MYB and R3-MYB protein families, respectively [17]. The corresponding gene for the activator is *NEGAN* (Nectar Guide Anthocyanin), and the gene for the inhibitor is *RTO* (Red Tongue). Figure 2.8 shows experimentally-created *Mimulus lewisii* mutants with varying levels of anthocyanin density.

Furthermore, *NEGAN* activates itself and *RTO*, *RTO* inhibits *NEGAN* function, and *RTO* is capable of intercellular (long-range) inhibition [17]. Therefore, the *NEGAN*-*RTO* system satisfies

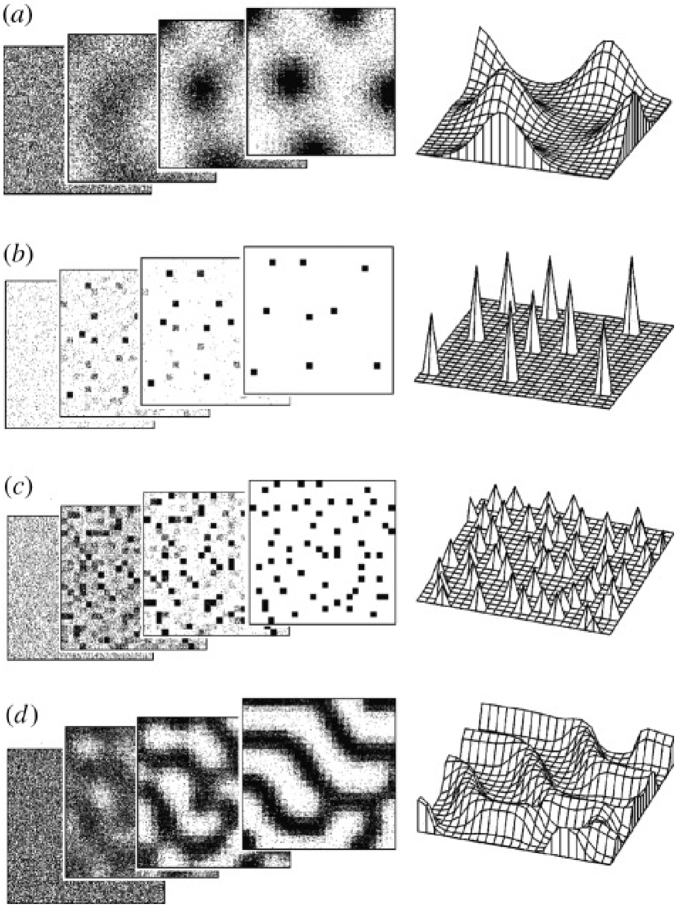


Figure 2.7: Examples of patterns produced by the Geirer-Meinhardt model. Figure from [53]

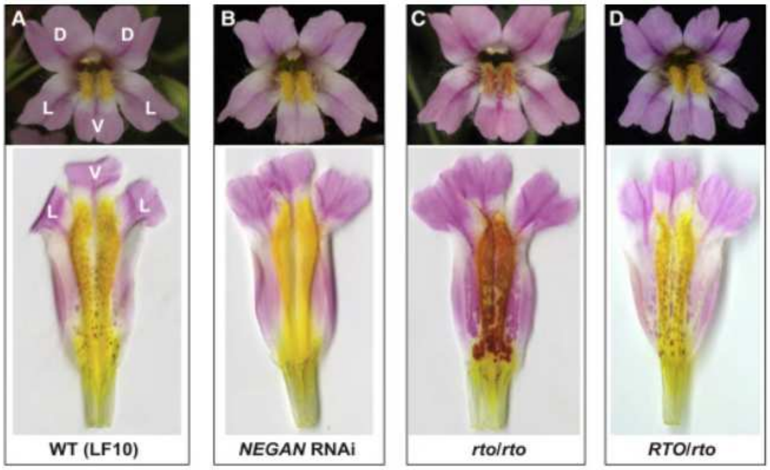


Figure 2.8: Wild type and experimentally-produced mutant *Mimulus lewisii* anthocyanin patterns. From left to right: (A) Wild-type, (B) Absence of anthocyanins due to *NEGAN* downregulation, (C) Anthocyanin overexpression from homozygous *rto* alleles, and (D) F1 plant with intermediate anthocyanin production due to incomplete dominance of *rto*. Figure from [17]

the requirements for the Gierer-Meinhardt model. Ding et al. proposed a modified version of the equations in (2.14):

$$\begin{aligned}\frac{\partial A}{\partial t} &= D_A \Delta A + G_A \frac{A^2 + A_0}{I + K} - U_A A \\ \frac{\partial I}{\partial t} &= D_I \Delta I + G_I A^2 + I_0 - U_I I.\end{aligned} \tag{2.15}$$

Here, the residual $K = 0.001$ prevents the denominator of $\frac{A^2+A_0}{I+K}$ from becoming 0, which allows the initial concentration of the inhibitor to be 0.

The Gierer-Meinhardt model produces patterns that look very similar to real *M. lewisii* spot patterns. We used the same parameters used by Ding and colleagues to generate simulated flower patterns: Diffusion constants $D_A = 0.01$ and $D_I = 0.5$, degradation rate constant $U_A = 0.03$, background production rate constants $A_0 = 0.01$ and $I_0 = 0$, and activation potency rate constants $G_A = 0.08$ and $G_I = 0.12$. The degradation constant of the inhibitor, U_I, was varied to produce different degrees of spot patterning. The initial activator concentration was 1M, while the initial inhibitor concentration was 0M.

Numerical simulations were done using a Fourier spectral method with periodic boundary conditions and fourth-order Runge Kutta method, given in [54]. The time scale of the simulations was chosen to be $t = 1500$ ($1.5 * 10^4$ iterations with a time step of $dt = 0.1$) because the patterns no longer evolved after this time. Our spatial domain was $0 \leq x, y \leq 50$, with a spatial step of $\frac{50}{256}$ and grid size 256x256. An example of the activator concentration surface for $U_I = 0.15$ is shown in Figure 2.9. This combination of parameters produces scattered arrangements of spots with smooth peaks, similar to the pattern seen in panel (a) of Figure 2.7.

The 2-dimensional images were produced by taking a level set at $z = 3$, which was also used by Ding and colleagues. Points with an activator concentration greater than 3 were assigned a value of 1, and the rest were assigned a value of 0. We colored the cells with value 1 red and the cells with value 0 yellow to produce images that resemble *M. lewisii* spots. Furthermore, we restricted our spatial grid to a size of 75×150, which gave us a narrow domain similar to the spotted area of the flower petals.

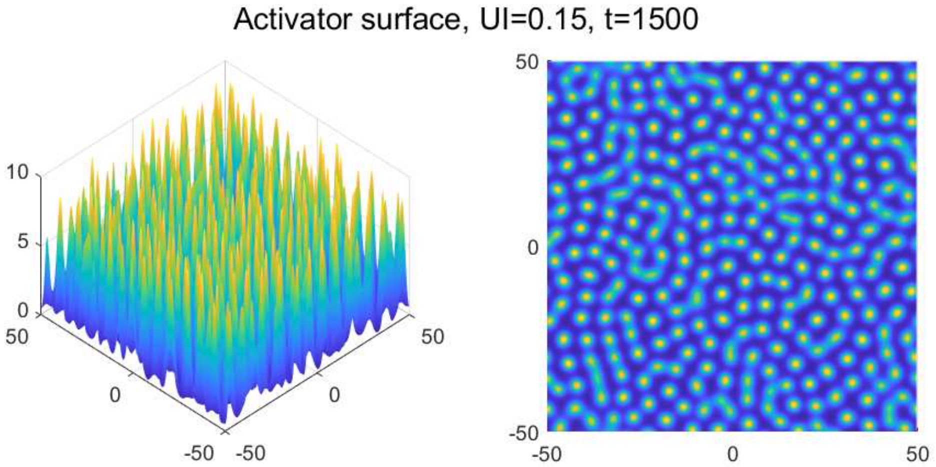


Figure 2.9: The activator concentration surface obtained using the Gierer-Meinhardt model.

A comparison between the spotted area of a *M. lewisii* petal and a simulated pattern is shown in Figure 2.10. Ding et al. visually matched this petal to $U_I = 0.06$ in the model, so we also chose $U_I = 0.06$ for this comparison.

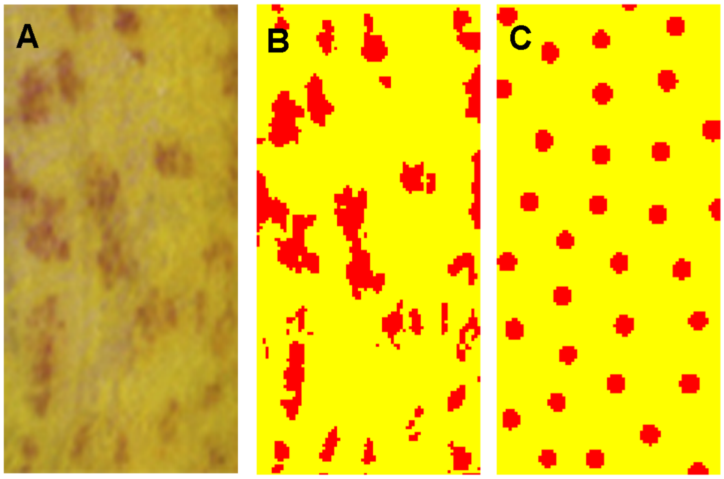


Figure 2.10: (A) A *M. lewisii* petal, (B) the same petal binarized in MATLAB with a threshold parameter of 0.53, (C) a simulated pattern with $U_I = 0.06$, restricted to a grid size of 75x150. Petal photo from [17].

In the following chapter, we will further discuss the role of the parameter U_I in spot formation, then introduce a method to classify these flower patterns by their U_I values using persistent homology.

Chapter 3

Pattern classification

3.1 Motivation

Our goal is to classify *Mimulus lewisii* petal patterns using their U_I values as the categories. Specifically, we seek to use the classifier on real data. We also have the broader goal of developing a classification method that can be used with petal patterns and their corresponding mathematical models in other flower species, in addition to other 2-D pattern-forming systems.

3.2 Pattern dependence on inhibitor degradation rate

Ding and colleagues modified the value of U_I, the degradation rate of the RTO inhibitor, in their modified Gierer-Meinhardt model (2.15) and found that spot size and density increased as U_I increased. We used U_I values in a range from 0.03 to 0.18 with step size 0.01 to generate the patterns we will use for classification (Figure 3.1). Furthermore, different values of U_I produce patterns that resemble *M. lewisii* with varying degrees of RNAi knockdown targeting the *RTO* gene [17]. Thus, the inhibitor degradation rate is a strong component in different degrees of spot expansion (Figure 3.2).

Persistent homology is the basis for our training data, so we also examined the persistence diagrams for each value of U_I. The persistent homology of the patterns varies with the value of U_I as well (Figure 3.3). Since variations in U_I cause dramatic changes in both the pattern images and their persistent homology, we determined that it would be feasible to classify the patterns based on their U_I values.

In the next sections, we will discuss persistent homology and its uses in machine learning in more detail.

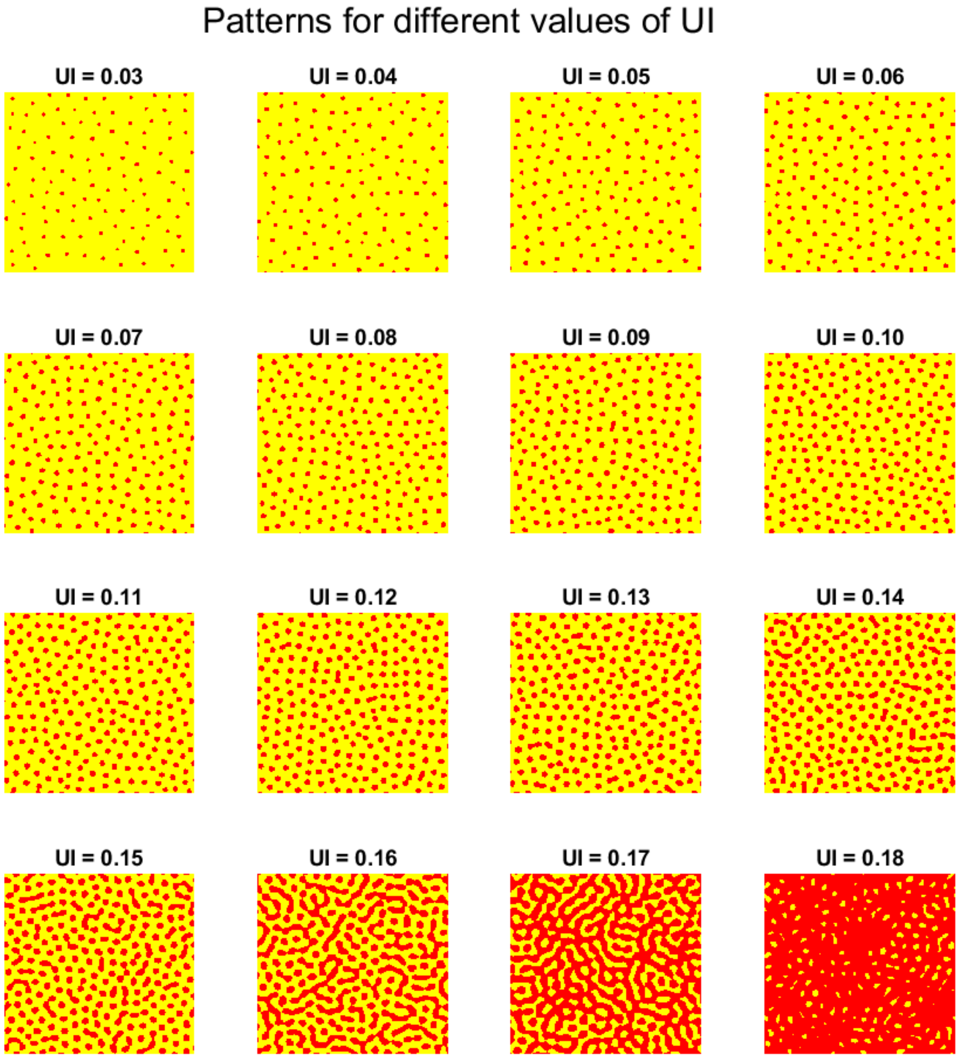


Figure 3.1: Patterns for values of U_I from 0.03 to 0.18, in increments of 0.01.

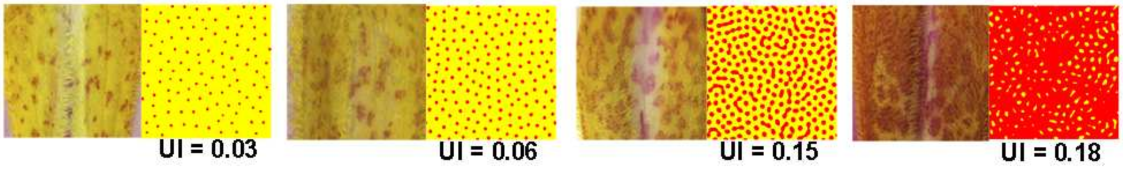


Figure 3.2: Simulated patterns that resemble incremental degrees of RNAi on the *RTO* gene in *M. lewisii*. Petal photos and parameter choices from [17].

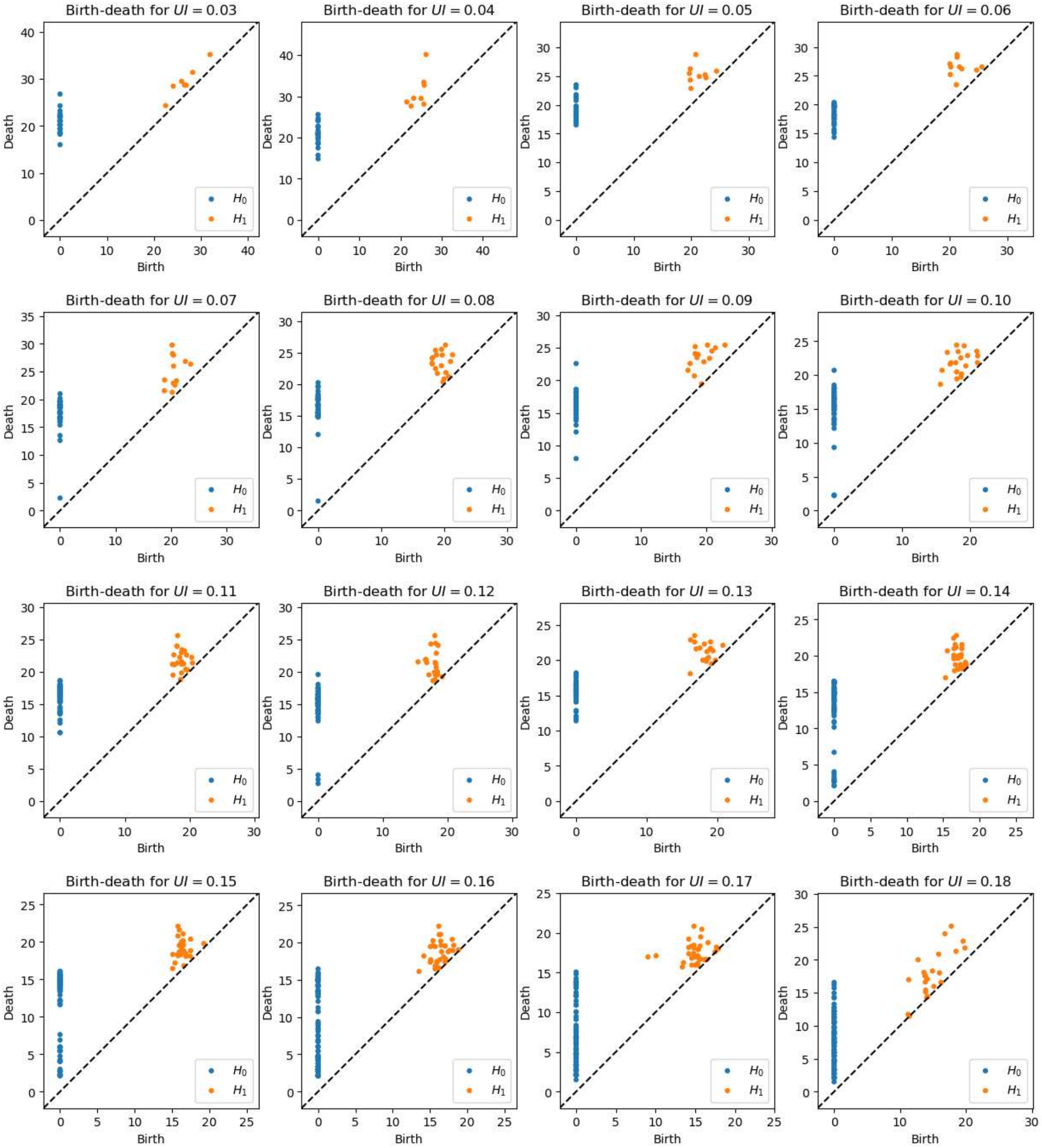


Figure 3.3: Persistence diagrams (H_0 in blue, H_1 in orange) for values of U_I from 0.03 to 0.18 in increments of 0.01.

3.3 Persistent homology

Persistent homology is a powerful computational topology tool that tracks the appearance and disappearance of features in a sequence of topological spaces. This information can be represented in a persistence diagram. Persistence diagrams are also referred to as birth-death diagrams because they are a collection of points (x, y), with each point representing a feature that appears (is born) at the scale x and disappears (dies) at scale y [55]. Generally, we consider features with longer lifespans (more "persistence") to be more important descriptors of the data.

For points in the plane, there are two topological features, namely connected components and holes. We illustrate persistent homology with the example set Ω in Figure 3.4, panels (a-c). The set Ω is a hexagonal lattice in which two points have been perturbed and one removed. These alterations introduce "holes" of different sizes into the pattern. Persistent homology allows us to characterize these holes. Suppose we place disks of radius r centered at each point in Ω. We connect any two points in Ω by an edge if the disks centered at these points enclose each other's centers. In addition, if three disks contain a triangle of edges, then that triangle is filled in by a face. The resulting simplicial complex, which consists of the points Ω, edges, and faces, is illustrated for a series of three progressively larger values of r, the connectivity parameter, in Figure 3.4 (a-c).

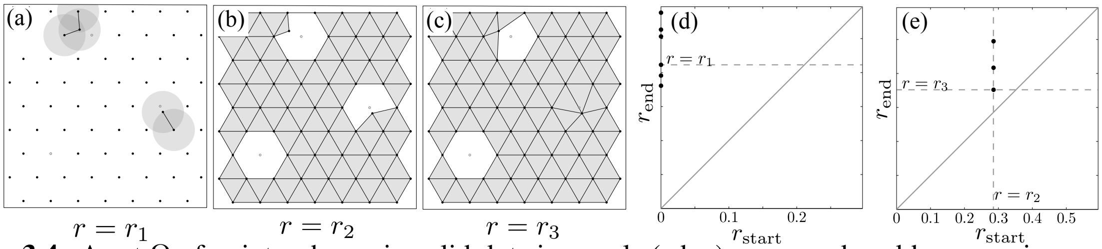


Figure 3.4: A set Ω of points, shown in solid dots in panels (a,b,c), was produced by removing one point and perturbing two points of a hexagonally arranged set of dots (open dots designate the original positions of dots that were altered). As the connectivity parameter r increases, edges and faces are added to form a sequence of simplices, examples of which are shown in panels (a,b,c). Each point (r_{start}, r_{end}) in the H_0 persistence diagram of panel (d) designates a value $r = r_{\text{start}}$ at which some connected component forms and the value $r = r_{\text{end}}$ of r at which that component ends by joining a connected component with a smaller value of r_{birth}. Similarly, the H_1 persistence diagram of panel (e) illustrates the values r_{start} and r_{end} of the connectivity parameter at which holes form and are filled in by faces, respectively.

For any value of r, we can count the number of holes in the simplicial complex, where a hole is defined to be a closed loop of edges that is not filled in by a face. There are, respectively, 0, 3, and 2 holes in the simplicial complexes of the first three panels of Figure 3.4. As we increase r, for any hole, there is a value r_{start} of the connectivity parameter at which the hole first appears and a value r_{end} at which it is filled in by a face. Each point in the persistence diagram shown in Figure

3.4 panel (e) is located at the coordinates $(r_{\text{start}}, r_{\text{end}})$ corresponding to a hole. This is called the H_1 persistence diagram.

The H_0 persistence diagram (Figure 3.4, panel (d)) is constructed similarly for components. Each point in the H_0 persistence diagram is located at the coordinates $(r_{\text{start}}, r_{\text{end}})$ corresponding to values of r at which a connected component forms and then merges with another component. Because our data will always have at least one component, the H_0 diagram contains a point of infinite persistence corresponding to the fully-connected simplical complex. We have omitted this point in our diagrams.

3.4 Persistence images

The space of persistence diagrams can be equipped with a metric, usually the Wasserstein or Bottleneck metric. These metrics are stable with respect to noise in the original data, so machine learning techniques can be used in conjunction with them. However, this approach is insufficient for some machine learning tools, and use of the metrics is very computationally expensive [56].

We solve this problem by using persistence images, which are stable vector representations of persistence diagrams [55]. The "pipeline" from persistence diagram to persistence image is summarized in Figure 3.5.

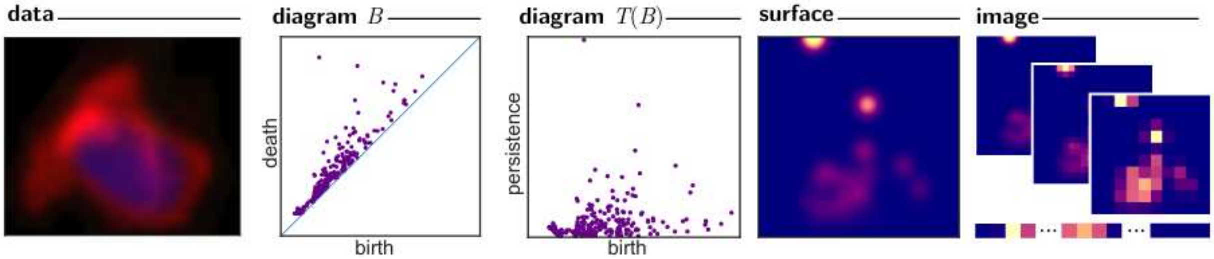


Figure 3.5: Pipeline to transform data into a persistence image. Figure from [55]

Let B be a persistence diagram (the set of points (x, y)). Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear tranformation $(x, y) \mapsto (x, y - x)$. Then, $T(B)$ transforms B so that the line $y = x$ becomes the horizontal axis.

Now, let $\phi_\mu : \mathbb{R}^2 \to \mathbb{R}$ be a differentiable probability distribution with mean $\mu = (\mu_x, \mu_y) \in \mathbb{R}^2$ and variance σ^2. We use $\phi_\mu = g_\mu$, where g_μ is the normal/Gaussian distribution with probabilty density

$$g_\mu(x, y) = \frac{1}{2\pi\sigma^2}\text{exp}\left(-\frac{(x-\mu_x)^2 + (y-\mu_y)^2}{2\sigma^2}\right) \tag{3.1}$$

We also require a weighting function $f : \mathbb{R}^2 \to \mathbb{R}$. The weighting function must be 0 along the horizontal axis, continuous, and piecewise differentiable. The weighting function should also be nondecreasing in y so it weights points with higher persistence more heavily. We use the weighting function used by Adams and colleagues in [55]. Let $b > 0$ and define $w_b(t)$ as follows:

$$w_b(t) = \begin{cases} 0 & t \leq 0 \\ \dfrac{t}{b} & 0 < t < b \\ 1 & t \geq b \end{cases} \tag{3.2}$$

We define $f(x, y) = w_b(y)$, with b defined to be the persistence value of the most persistent feature.

The above components are combined to convert the persistence diagram B into a persistence image. First, we create a persistence surface $\rho_B : \mathbb{R}^2 \to \mathbb{R}$ by taking the Gaussian distribution centered at each point in $T(B)$, then multiplying it by the weight of that point:

$$\rho_B(z) = \sum_{u \in T(B)} f(u)\phi_u(z) \tag{3.3}$$

The assumptions we made for the weighting function f ensure the stability of the transformation of the persistence diagram into a persistence surface. The surface ρ_B is reduced to a finite-dimensional vector by discretizing the plane: we fix a grid of n boxes or pixels in the plane, then integrate ρ_B over each box. The persistence image I of B is defined as follows for $1 \leq p \leq n$:

$$I(\rho_B)_p = \int\int_p \rho_B dy dx \tag{3.4}$$

In our persistence images (images for H_1 are shown in Figure 3.6), we used the Gaussian distribution shown in (3.1), the weighting function defined in (3.2), and a pixel size of 1.

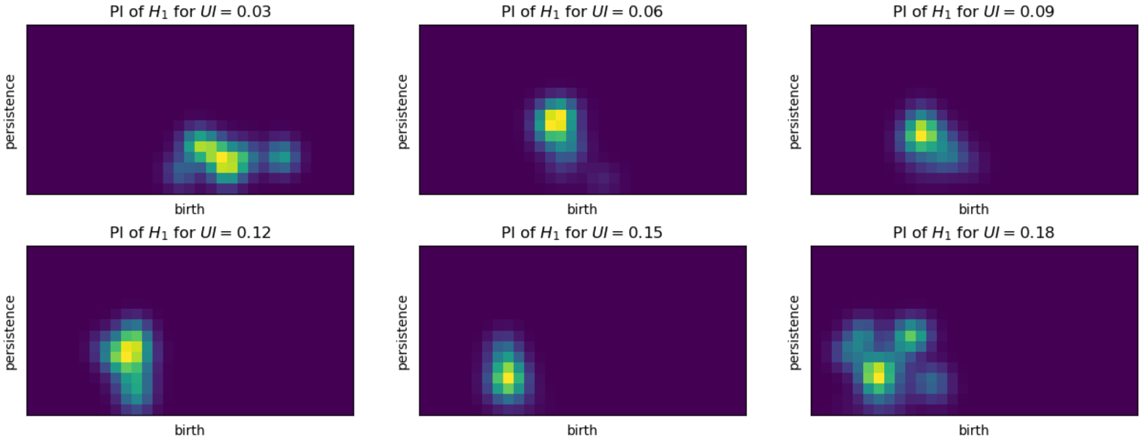


Figure 3.6: Examples of H_1 persistence images from simulated spot patterns. Compare to the persistence diagrams seen in Figure 3.3.

3.5 Support vector classification

We will now summarize the machine learning approach that we will use to classify our pattern data. The goal of classification is to assign input vectors **x** to one of k discrete categories, $\{C_i\}_{i=1}^k$. In our case, the input vectors are the persistence images, and the categories are the value of U_I that should be used to generate the pattern. The categories are taken to be disjoint: that is, each input vector is only assigned to one category. Therefore, the input space can be divided into "decision regions" with boundaries called "decision boundaries" [57].

Machine learning algorithms are tasked with determining the location of the decision boundaries. There are many such algorithms, but we will use the linear Support Vector Machine (SVM). If we have n-dimensional training data, the decision boundary will be an $n-1$-dimensional hyperplane. SVM places the hyperplane in a way that maximizes the distance (called the margin) from the hyperplane to the nearest training points of each category. The nearest training points are referred to as the support vectors (Figure 3.7) [58]. Thus, only a small amount of the training data is needed to construct the hyperplanes; points outside of the margin can be moved around without affecting the solution to the optimization problem [57].

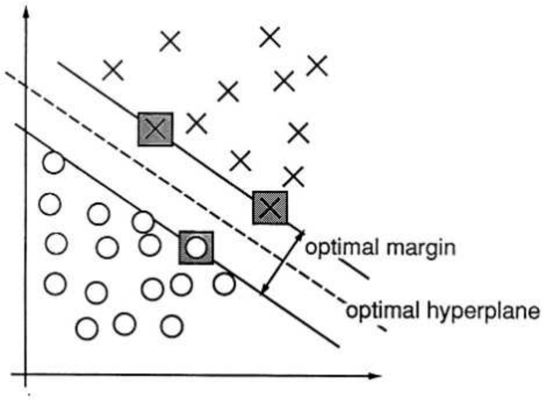


Figure 3.7: An example of a binary SVM problem in 2 dimensions. The support vectors are marked with grey squares. Figure from [58]

SVM is formulated as a binary classification algorithm, but it has been extended for multiclass problems. Suppose we have a classification problem for k categories. There is the "one-vs-rest" method, which trains k binary SVMs: for the ith SVM, category i is separated from the rest of the categories. The second approach is the "one-vs-one" method, where $\frac{k(k-1)}{2}$ binary SVMs are trained to separate pairwise combinations of the categories [57]. Our classifier uses the "one-vs-one" method.

3.6 Classification results

We used MATLAB to generate our point data. For classification, we used the following Python packages: Ripser [59] to generate persistence diagrams, Persim to generate persistence images, and Scikit-Learn [60] for the classifier. We generated 120 simulations with random initial conditions for each value of U_I from 0.03 to 0.18 in increments of 0.01. In the simulations where the 75×150 grid size was used, we used 3 samples from different areas for each pattern, bringing the total number of datasets up to 360.

The data were randomly assigned to the training and test sets with a split of 80% training data and 20% test data. We ensured consistency between methods by using a seed of 2. Persistence images of different dimensions can be used alone or concatenated with each other [55], so we trained classifiers using H_0, H_1, and H_0 and H_1 concatenated.

Accuracy is calculated as the ratio of correct classifications (classifications with $U_I^{predicted} = U_I^{actual}$) to the total number of patterns in the test set. Our increment between U_I values is very small, resulting in subtle differences in the spot locations of adjacent patterns. Therefore, we also include a quantity that we refer to as "off-diagonal accuracy," where we sum (classifications with $U_I^{predicted} = U_I^{actual}$) + (misclassifications where $U_I^{predicted} = U_I^{actual} \pm 0.01$) before dividing by the total number of test patterns.

3.6.1 Classification using 3-D data

First, we used the 3-dimensional activator concentration surfaces from the Gierer-Meinhardt model (shown in Figure 2.9) to train the classifier. The surfaces were converted into point data by taking the x- and y-values of their relative maxima, which correspond to the locations of the surface's peaks, using the extrema2 function [61].

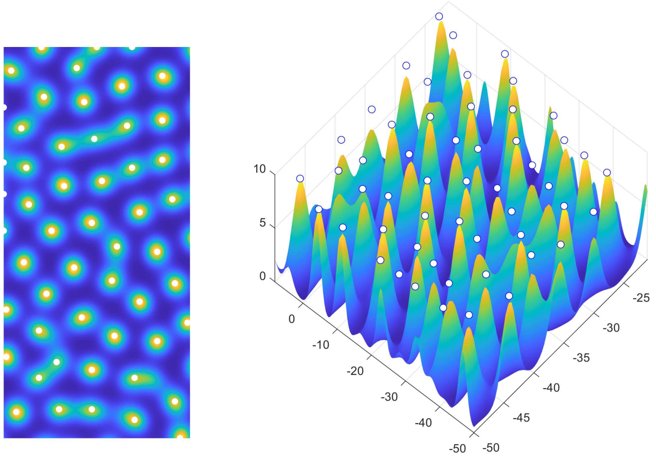
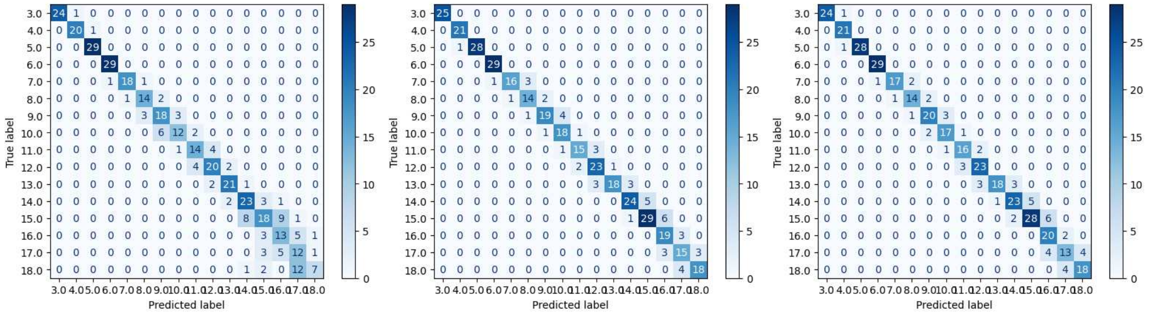


Figure 3.8: The locations of the relative maxima overlaid on a surface with $U_I = 0.15$.

We began by using the extrema of the entire 256×256 grid size. The confusion matrices are shown in Figure 3.9. Classification with H_0 was 76.0% accurate, H_1 was was 86.2% accurate, and H_0 and H_1 was 85.7% accurate. Using both H_0 and H_1 showed no improvement over H_1 alone. Off-diagonal accuracies were 85.9% for H_0, 100% for H_1, and 100% for H_0 and H_1.

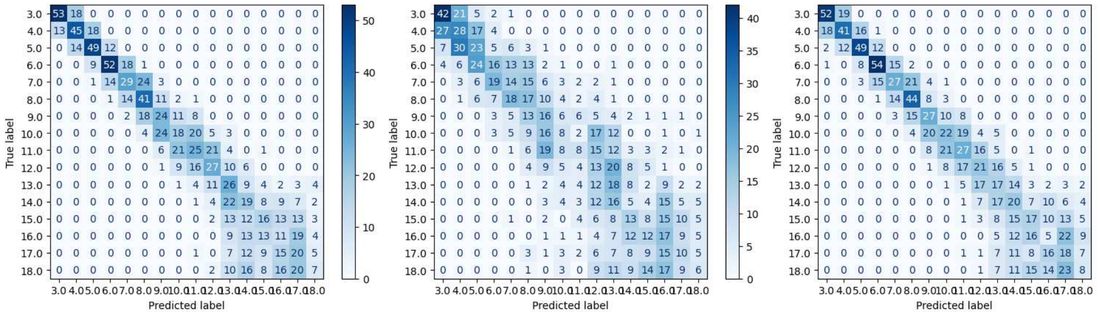


(a) H_0, 76.0% (85.9%) accurate **(b)** H_1, 86.2% (100%) accurate **(c)** H_0 and H_1, 85.7% (100%) accurate

Figure 3.9: Confusion matrices using the relative maxima of the original activator concentration surfaces, using training data with domain 256×256. Percentages in parentheses are off-diagonal accuracies.

Classification results were very good for the full grid size, but we also tried restricting the grid to 75×150 to match the real petal spot domain (see Figure 2.10). Confusion matrices are shown in Figure 3.10. Classification with H_0 was 40.1% accurate, H_1 was 21.61% accurate, and H_0 and H_1 was 39.0% accurate. Off-diagonal accuracies were 79.3% for H_0, 56.4% for H_1, and 78.6% for H_0 and H_1.

Classification accuracy decreased dramatically when we restricted the grid size. Additionally, this method cannot be extrapolated to real data because it requires the 3-D activator concentration surface, which we can only obtain through simulations. However, we found that the topology of patterns with different U_I values is distinct enough to classify using the locations of the spots, so we turned our attention to the 2-D pattern images.



(a) H_0, 40.1% (79.3%) accurate **(b)** H_1, 21.61% (56.4%) accurate **(c)** H_0 and H_1, 39.0% (78.6%) accurate

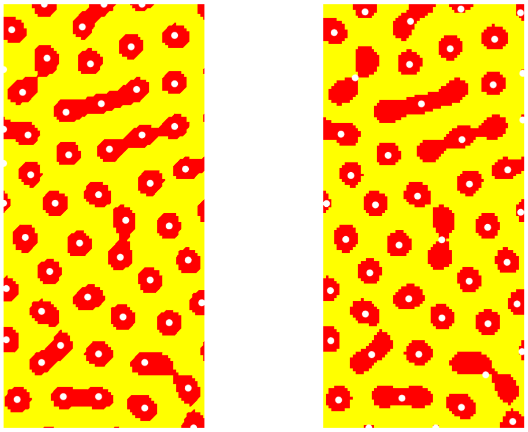
Figure 3.10: Confusion matrices using the relative maxima of the original activator concentration surfaces, using training data with domain 75×150. Percentages in parentheses are off-diagonal accuracies.

3.6.2 Classification using 2-D data: Center of mass

Next, we converted the level sets seen in Figure 3.1 into point data by taking the center of mass of each cluster of 1s (the red spots). One limitation of this method was immediately apparent: we could not use it to generate data for the images corresponding to U_I values of 0.17 and 0.18. These images had so much red that only one center of mass could be generated, which provides insufficient data for persistent homology.

In the rest of the patterns, taking centers of mass resulted in only one point for each red area. As a result, these points only captured the density of the spots and not their size. The center of mass points were especially poor descriptors of the patterns with higher U_I values that led to merged spots. These larger spots appeared in the point data only once, despite their larger area in the pattern, as shown in Figure 3.11.

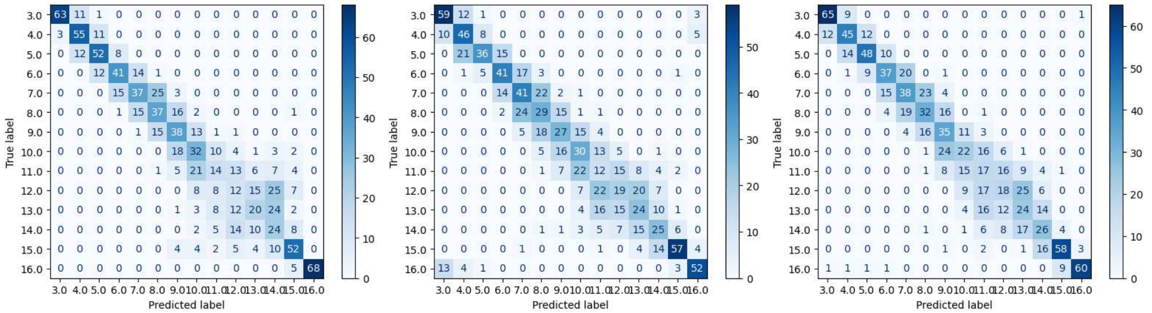
Despite these limitations, the classification had fairly good accuracy when tested on simulated data. Confusion matrices are shown in Figure 3.12. Classification with H_0 was 54.0% accurate, H_1 was 49.4% accurate, and H_0 and H_1 was 52.1% accurate. Off-diagonal accuracies were 86.2% for H_0, 86.2% for H_1, and 89.1% for H_0 and H_1. The lack of information for the shapes of larger spots is apparent as the value of U_I increases.



(a) Extrema of the original surface

(b) Centers of mass of the level set

Figure 3.11: A comparison between the extrema of the original surface and the centers of mass of the level set for the same pattern with $U_I = 0.15$.



(a) H_0, 54.0% (86.2%) accurate

(b) H_1, 49.4% accurate (86.2%)

(c) H_0 and H_1, 52.1% (89.1%) accurate

Figure 3.12: Confusion matrices for classification with the centers of mass of the 2-D patterns. Percentages in parentheses are off-diagonal accuracies.

Spots in real petals are often much less uniform in organization and shape than the simulated data that the classifier was trained on, so we looked for a data processing method that could better capture the information in the patterns.

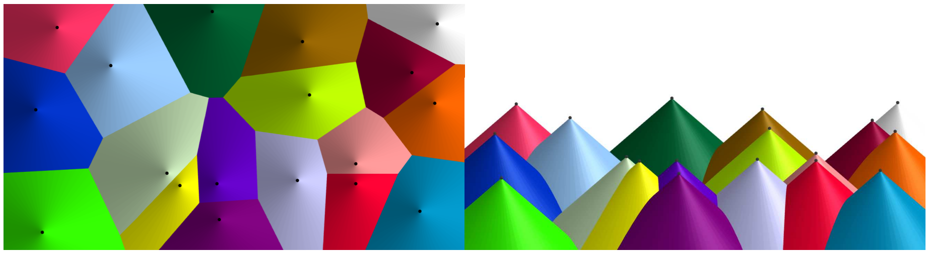
3.6.3 Classification using 2-D data: Voronoi surfaces

The center of mass method resulted in the loss of a large amount of topological information, so we moved on to developing a method to extract 3-dimensional information from the level sets.

We constructed 3-dimensional surfaces by using an algorithm similar to the one used to generate 3-dimensional Voronoi diagrams, also referred to as Voronoi cones or mountains.

A 2-dimensional Voronoi diagram is constructed using a set of points, referred to as generators or seeds, $\{p_i\}_{i=1}^{n}$. The plane is divided into regions, $\{R_i\}_{i=1}^{n}$, where p_i is inside R_i for each i. The distance from every point in the plane to the generators is calculated with the l^2 distance, but other distance functions can be used. All the points in region R_k ($1 \leq k \leq n$) are closer to the generator p_k than they are to any other generator p_i (where $i \neq k$). The points that form the boundaries of each region are equally close to the two nearest generators.

Voronoi diagrams are converted into Voronoi cones by setting the z value of each point equal to its distance from the generator of its region. This surface may be reflected across the xy-plane so the generators form the peaks of the cones. In MATLAB, we accomplished this by using the voronoi mountains function [62].



(a) Voronoi diagram **(b)** Voronoi cones

Figure 3.13: An example of a Voronoi diagram and its corresponding Voronoi cones. Images generated with [63].

We used the indices of each red cell from the level set matrix to create our set of generating points. We generated a Voronoi surface V_1 with this set. However, every point in a spot was treated as a generator, preventing us from obtaining defined peaks. We solved this problem by creating a second Voronoi surface V_2 using the indices of each yellow cell as generators, which gives us the distances from the red points to the nearest yellow area. We combined these surfaces by replacing the red areas in V_1 with the negative z-values from V_2, which gave us smooth peaks in the center of each red spot. Finally, we reflected the surface over the xy-plane so the red areas would form peaks as they did in the original surface.

For $U_I = 0.03$ and $U_I = 0.04$, the spots were so small that this method did not yield point data that was accurate to the patterns. For these sets, we found the extrema of V_2 (the surface using the yellow points as generators) directly.

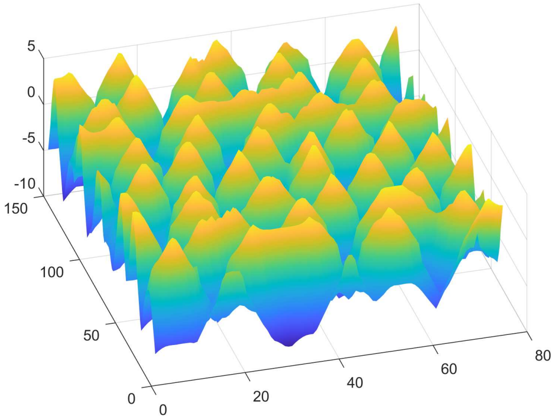
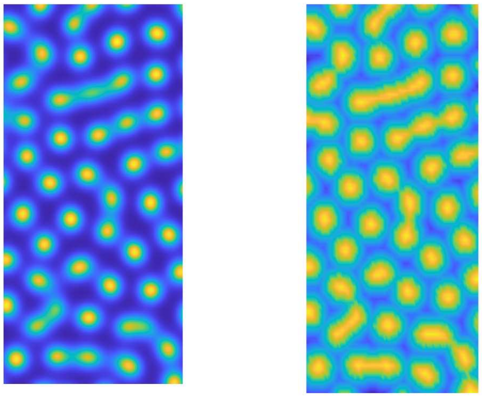


Figure 3.14: An example of the Voronoi surface for a pattern with $U_I = 0.15$.

This method recovered the behavior of the original surface very well. A comparison between the original surface and its Voronoi reconstruction from the level set is shown in Figure 3.15.

We obtained point data by taking the x- and y-values of the relative maxima (as we did for the original surfaces). A comparison of the points generated from the original surface and the Voronoi surface is shown in Figure 3.16. Notably, the Voronoi surface had more relative maxima within the larger spots than the original surface did.

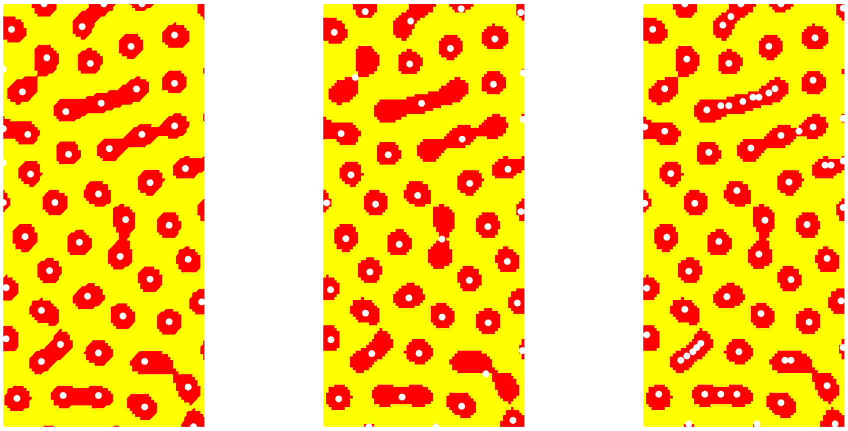
Confusion matrices are shown in Figure 3.17. Classification with H_0 was 66.8% accurate, H_1 was 60.2% accurate, and H_0 and H_1 was 64.8% accurate. The Voronoi-surface approach thus acheived the best accuracy of the data processing methods. Off-diagonal accuracies were 97.1% for H_0, 95.0% for H_1, and 97.5% for H_0 and H_1. Therefore, we also obtained the best off-diagonal accuracy using the Voronoi-surface method. Classification was most successful for the U_I values on the highest end of the range.



(a) The original surface

(b) The Voronoi surface

Figure 3.15: A comparison of the original activator concentration and Voronoi surfaces for the same simulated pattern with $U_I = 0.15$.



(a) Extrema of the original surface

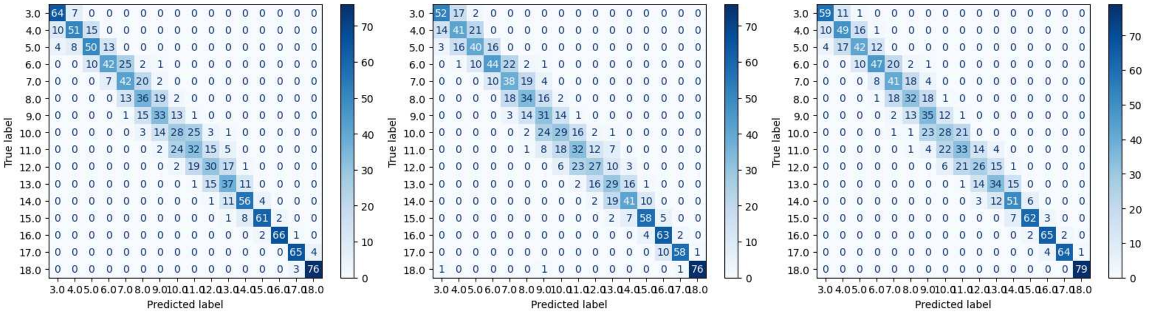
(b) Centers of mass of the level set

(c) Extrema of the Voronoi surface

Figure 3.16: A comparison of the points generated by the data processing methods we used for the same simulated pattern with $U_I = 0.15$.

3.6.4 Choice of level set for Voronoi surface

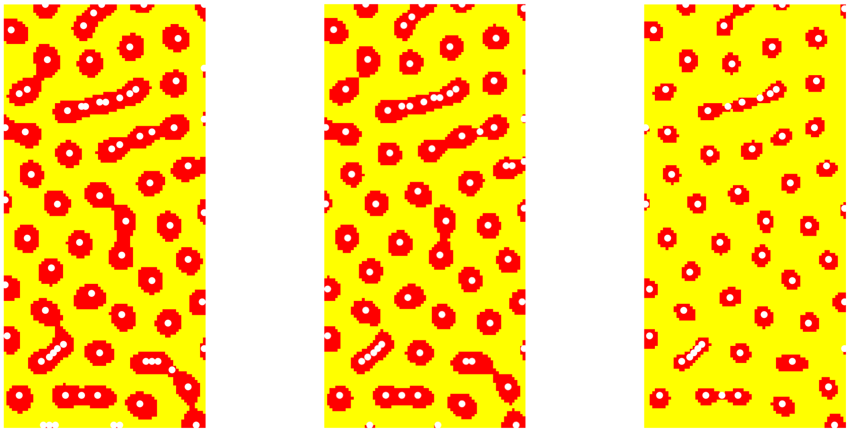
The choice $z = 3$ for the level set was arbitrary and based on the resemblance of the simulated images to real patterns. We generated additional patterns using different thresholding values to determine whether the topological features are preserved at different level sets.



(a) H_0, 66.8% (97.1%) accurate **(b)** H_1, 60.2% (95.0%) accurate **(c)** H_0 and H_1, 64.8% (97.3%) accurate

Figure 3.17: Confusion matrices for classification with extrema of the Voronoi surfaces. Percentages in parentheses are off-diagonal accuracies.

We found that $z = 4.5$ and $z = 2.65$ were the highest and lowest level sets, respectively, that preserved the patterns for all U_I values and gave persistence diagrams with enough information for classification. A comparison between patterns thresholded with $z = 2.65$, $z = 3$, and $z = 4.5$ is shown in Figure 3.18.



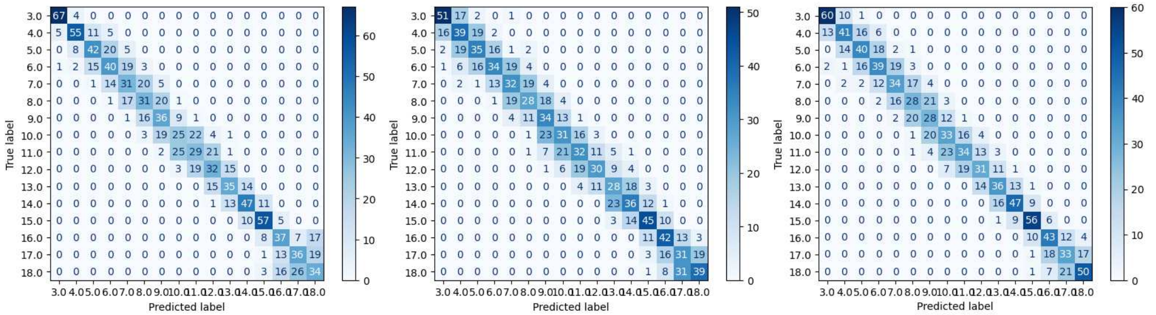
(a) Level set at $z = 2.65$ **(b)** Level set at $z = 3$ **(c)** Level set at $z = 4.5$

Figure 3.18: A comparison of the 2-D pattern images and the maxima of their Voronoi surface reconstructions for different level sets. The pattern has $U_I = 0.15$.

Confusion matrices are shown in Figure 3.19 for $z = 4.5$ and 3.20 for $z = 2.65$. Classification using the $z = 4.5$ level set with H_0 was 55.0% accurate, H_1 was 49.2% accurate, and H_0 and H_1

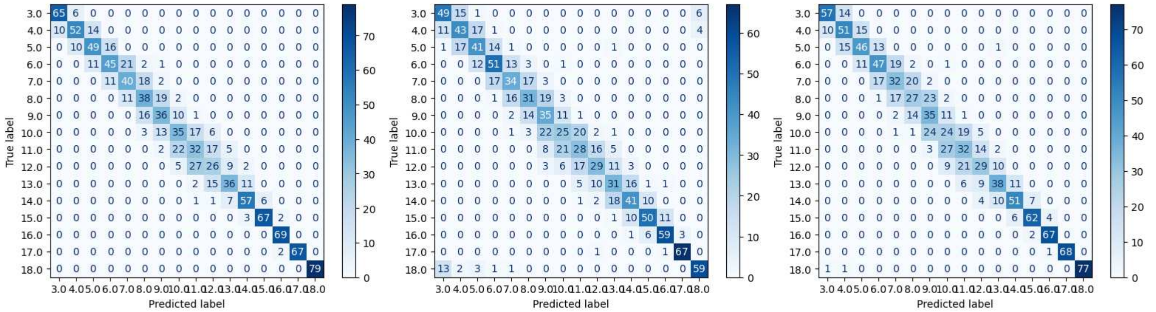
was 54.9% accurate. Off-diagonal accuracies were 93.2% for H_0, 91.1% for H_1, and 94.1% for H_0 and H_1.

Classification with the $z = 2.65$ level set with H_0 was 74.3% accurate, H_1 was 97.9% accurate, and H_0 and H_1 was 97.5% accurate. Off-diagonal accuracies were 97.0% for H_0, 91.8% for H_1, and 96.1% for H_0 and H_1.



(a) H_0, 55.0% (97%) accurate **(b)** H_1, 49.2% (91.1%) accurate **(c)** H_0 and H_1, 54.9% (94.1%) accurate

Figure 3.19: Confusion matrices for classification with extrema of the Voronoi surfaces using $z = 4.5$. Percentages in parentheses are off-diagonal accuracies.



(a) H_0, 68.8% (97.0%) accurate **(b)** H_1, 58.4% (91.8%) accurate **(c)** H_0 and H_1, 64.5% (96.1%) accurate

Figure 3.20: Confusion matrices for classification with Voronoi surfaces using $z = 2.65$. Percentages in parentheses are off-diagonal accuracies.

Both raw and off-diagonal accuracy decreased slightly from level set 3, but were still high. We tested the highest and lowest level sets that provided us with meaningful topological information and found that classification accuracy is relatively unchanged. Therefore, we conclude that the unique surface topology of each pattern is preserved between level sets.

3.6.5 Summary of classification accuracy

We summarize the classification accuracies for each method we tested in Table 3.1. We additionally provide computation times for each step of the process in Table 3.2. Times for data processing are given as an approximation for one dataset. Times for persistent homology and classification include all three dimensions we used for the entire data.

Table 3.1: A summary of accuracies for each classification method.

Method	H0	H1	H0 and H1
Surface maxima, 256×256	76.0	86.2	85.7
Off-diagonal accuracy	85.9	100	100
Surface maxima, 75×150	40.1	21.6	39.0
Off-diagonal accuracy	79.3	56.4	78.6
Centers of mass	54.0	49.4	52.1
Off-diagonal accuracy	86.2	86.2	89.1
Voronoi maxima, $z = 3$	66.8	60.2	64.8
Off-diagonal accuracy	97.1	95.0	97.3
Voronoi maxima, $z = 4.5$	55.0	49.2	54.9
Off-diagonal accuracy	93.2	91.1	94.1
Voronoi maxima, $z = 2.65$	68.8	58.4	64.5
Off-diagonal accuracy	97.0	91.8	96.1

Table 3.2: Computation times for each method (in seconds).

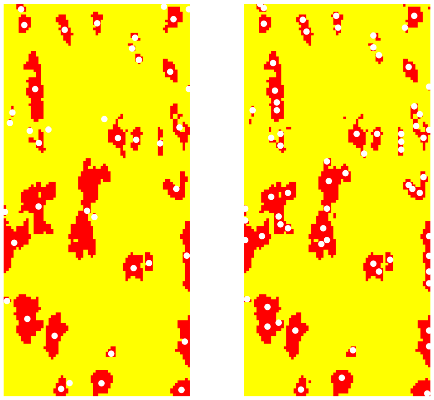
Method	Processing	PH	Training
Surface maxima, 256×256	0.15	189.5	1.27
Surface maxima, 75×150	0.06	18.5	36.6
Centers of mass	0.002	18.6	83.5
Voronoi surface	0.2	29.41	51.4

3.7 Petal data

We tested both of the 2-D classification methods with *M. lewisii* photos taken by Ding and colleagues [17]. Figure 3.2 lists the photos alongside the simulations that Ding and colleagues matched them with based on visual resemblance. We use these identifications to compare the results of machine learning with identifications done by humans.

We split the petals in half because the shadows from the groove in the center led to inaccurate thresholding. Thresholding was done in MATLAB, with parameters chosen manually for each image. We chose parameters so the thresholding captured as much spot information as possible without including noise from the photos.

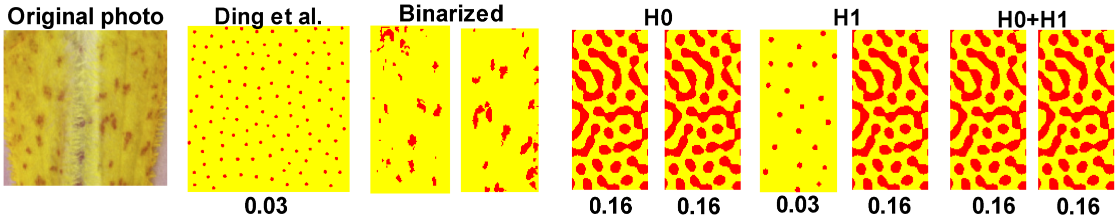
As previously mentioned, the center of mass method only captured information about spot density and not size or shape, which caused the point clouds for real patterns to be relatively poor descriptors of the data. We also could not classify the mostly red petal (Figure 3.23, panel (d)) with centers of mass because we could not train the classifier with patterns for $U_I = 0.17$ and 0.18. Conversely, the Voronoi method captured the pattern on a real flower better than the center of mass method did, due to its ability to capture the shape information in larger spots. A comparison between the centers of mass and Voronoi maxima for the same petal are shown in Figure 3.21.



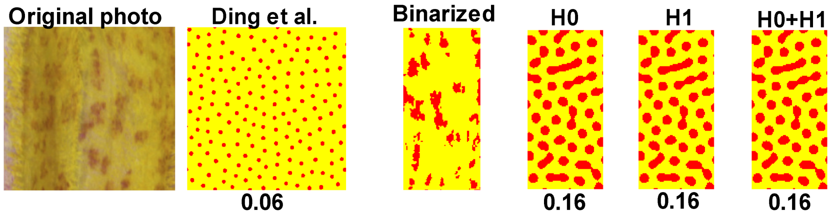
(a) Centers of mass **(b)** Voronoi maxima

Figure 3.21: Comparison of centers of mass and Voronoi maxima for the same petal.

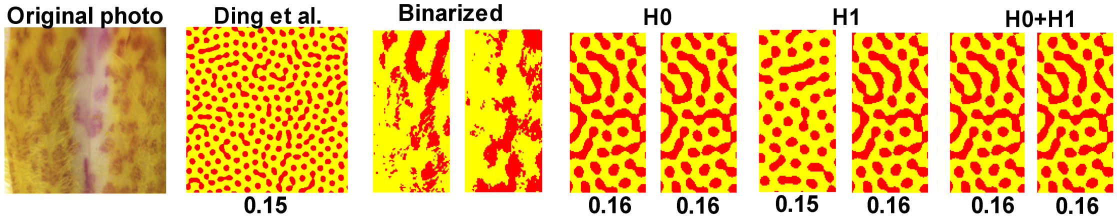
We found that using centers of mass almost entirely gave classifications of $U_I = 0.15$ and 0.16, regardless of the dimension of persistent homology (Figure 3.22). When we used the Voronoi surfaces, we obtained more plausible results for 2 of the petals (Figure 3.23, panels (a,d)). However, the classifier was less successful when presented with the petal with intermediate spotting (panel b,c), predicting $U_I = 0.18$ for both.



(a) $U_I = 0.03$

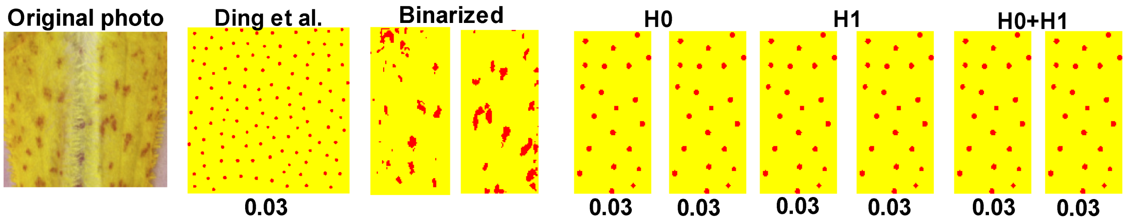


(b) $U_I = 0.06$

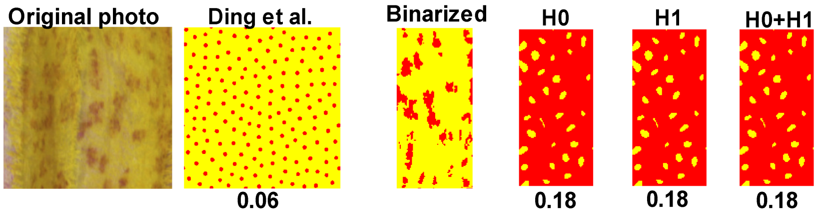


(c) $U_I = 0.15$

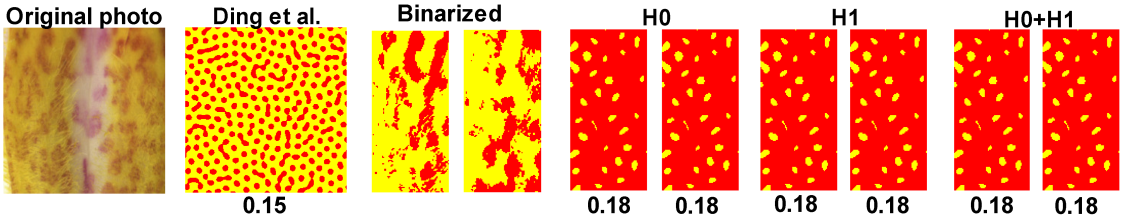
Figure 3.22: Results for real petal classification using centers of mass for the spots. Petals are labelled based on the parameter given to them in Ding et al.



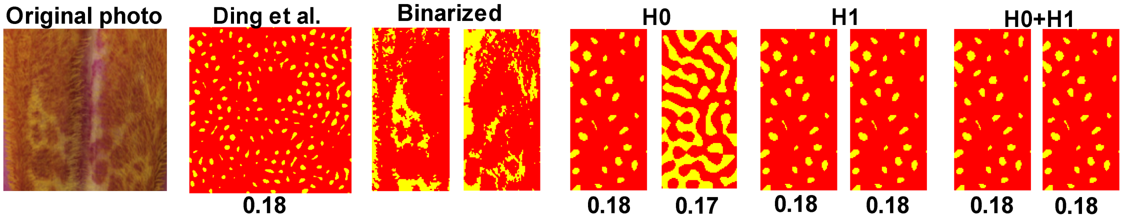
(a) $U_I = 0.03$



(b) $U_I = 0.06$



(c) $U_I = 0.15$



(d) $U_I = 0.18$

Figure 3.23: Results for real petal classification using Voronoi surfaces. Petals are labelled based on the parameter given to them in Ding et al.

Chapter 4

Conclusions and discussion

4.1 Conclusions

In this thesis, we found that training a classifier using persistent homology in conjunction with the Gierer-Meinhardt model yields enough information for us to successfully categorize flower pattern data. Patterns can be classified using persistent homology on points obtained directly from the surface given by the Gierer Meinhardt model (Section 3.6.1). However, we can also use persistent homology with 2-D pattern images, which gives us the ability to classify real flower patterns using simulated training data. Notably, classification is based solely on the shape and location of spots, rather than their radius.

We have presented a novel method for processing 2-D pattern data by using a modified Voronoi surface algorithm. This method is very successful with simulated training data, and we have shown that it can produce plausible classifications for real flower patterns. Furthermore, this was possible even with relatively limited information (a small pattern domain) and subtle differences between each pattern category. If we compare our results to the results for the other two methods, we find that a key advantage is its ability to capture long spots, as evidenced by its improvement in performance for higher U_I values.

4.2 Discussion

4.2.1 Role of H_0 and H_1

Regardless of the method we used to process the data, we found that the H_1 persistence images were consistently worse for classification than H_0. Furthermore, concatenating H_0 and H_1 improved accuracy compared to H_1 alone, but not H_0.

One hypothesis for this result is that H_0 persistent homology may contain more information about the shape of the spots than H_1. The sizes of the strings of points we see in longer spots (Figure 3.16), or their absence, are included in the H_0 persistence diagram.

However, when we take into account the off-diagonal classifications, H_0 and H_1 concatenated yields better results than either on their own, although H_0 alone still has better accuracy than H_1. This indicates that, while H_0 persistence images provided a better basis for accurate classification, adding H_1 to it added more precision in the misclassifications.

4.2.2 Category density and regression

Here, we took a relatively fine increment of 0.01 between U_I values. Because most misclassifications were for neighboring values of U_I, we infer that taking coarser increments would improve classification accuracy, while finer increments would worsen it.

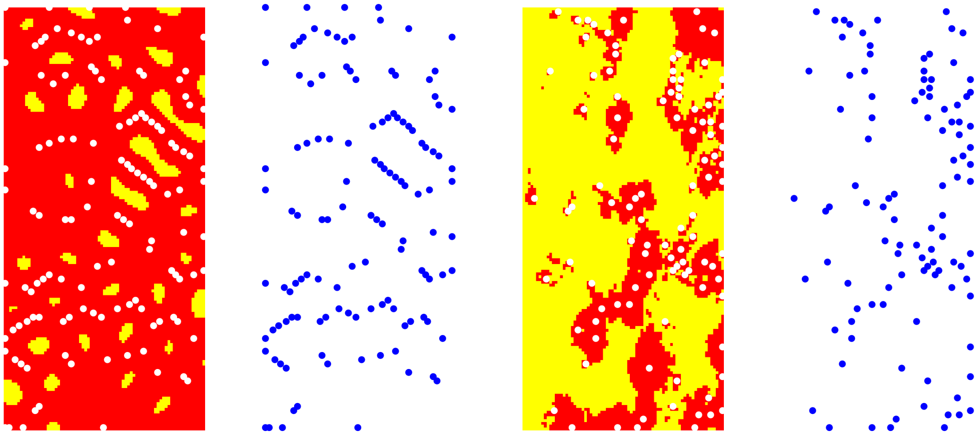
Rather than taking progressively finer increments between categories, future work could treat pattern classification as a regression problem, which would give us a continuous probability distribution for each value of U_I. This would allow us to characterize patterns more precisely without using increasingly small pattern categories.

4.2.3 Improving classification for real data

Classification for real data could be improved further by modifying the training data to better resemble real patterns.

While the simulated patterns produced by the Gierer-Meinhardt model bear a resemblance to *M. lewisii* spot patterns, their uniform spot density and spot sizes limit this similarity. This limits the accuracy of classification for petals with intermediate degrees of spotting because they often have spots with irregular shapes, sizes, and spatial position. Future work could modify the training data to better equip the classifier for such patterns.

One possibility is to randomly remove spots from the simulated patterns, which would expose the classifier to training data containing empty spaces similar to what we see in real *M. lewisii* petals. Another possibility would be to modify our algorithm for patterns where the "red" area covers most of the plane, as the points generated by our Voronoi surface do not account for all of the red in the pattern (Figure 4.1). This is a potential explanation for the results shown in Figure 3.23, panels (b) and (c).



(a) Points over a pattern with $U_I = 0.18$

(b) Scatter plot of the simulated points

(c) Points over a real petal

(d) Scatter plot of the petal points

Figure 4.1: A comparison of the points for a pattern with $U_I = 0.18$ to those for the petal seen in Figure 3.23, panel (c).

Additionally, one limitation of using binary images as inputs is that classification of photos is dependent on the quality of the photo and how it was thresholded (particularly because the thresholding was done at our discretion). Some of the petal photos that we used had inconsistencies in lighting, so a portion of the spots did not appear in the thresholded images. Generally, we also wish to test the classifier on more *M. lewisii* petal images with a wider range of spot patterns. Future development of the classifier should use better quality images in greater numbers to increase the quality of classification.

Finally, our classification method should be tested on other petal pattern formation models. If it can be used to classify 2-D patterns in other pattern systems, we could better discern the effects of visually-similar patterns on pollinator preference and learning. It would also provide insight into the significance of factors that underlie petal pattern development.

4.2.4 Beyond flower patterns

Although this thesis has focused on the use of our classification method for pigment patterns on flower petals, it has the potential for use in other pattern-forming systems because it only requires 2-D images as input data. One interesting example in ecology is dryland vegetation. Dryland vegetation appears inconsistently patchy from the ground, but aerial photography reveals orderly arrangements of plants resembling Turing's reaction diffusion patterns (Figure 4.2) [64].

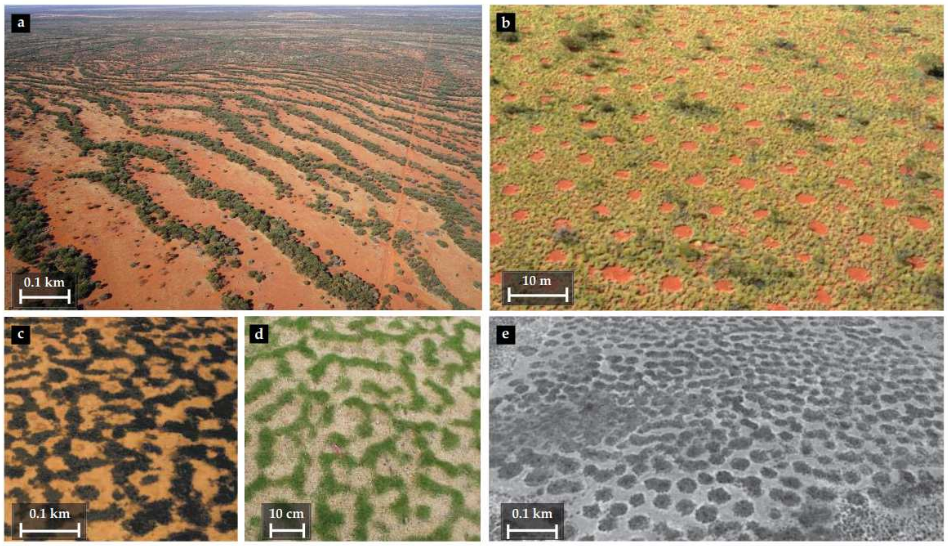


Figure 4.2: Large-scale reaction diffusion patterns formed by dryland vegetation. Figure from [64].

These patterns are the result of a water-biomass feedback loop caused by areas with denser vegetation drawing water away from its surroundings [64, 65]. Different mechanisms of water transport induce positive feedback loops that can act separately or together to produce pattern-forming instabilities; their relative importance changes with environmental conditions. Furthermore, this mechanism can be modelled using a system of partial differential equations, giving a quantitative basis for large-scale vegetation patterning [66].

If we view vegetation pattern formation as a function of precipitation, the results look remarkably similar to the patterns in Figure 2.7. Characterizing the patterns between multiple ecosystems could provide insight into the environmental conditions that drive them. In the face of a changing global climate, patterns could also serve as an indicator for unusual environmental conditions and a guideline for human intervention.

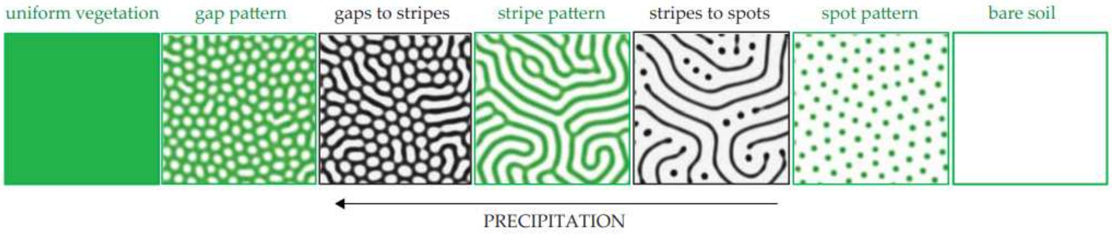


Figure 4.3: An illustration of vegetation pattern dependence on precipitation. Figure from [64].

Vegetation patterning is just one example of a potential application for 2-D pattern classification. Orderly pattern formation occurs in many more biological systems, and using machine learning to characterize their parameters would provide a better understanding of their effects and the mechanisms that underlie them [65].

Bibliography

[1] E. Grotewold. The genetics and biochemistry of floral pigments. *Annual Review of Plant Biology*, 57(1):761–780, 2006. PMID: 16669781.

[2] X. T. Zheng, Z. C. Yu, J. W. Tang, M. L. Cai, Y. L. Chen, C. W. Yang, W.S. Chow, and C. L. Peng. The major photoprotective role of anthocyanins in leaves of arabidopsis thaliana under long-term high light treatment: antioxidant or light attenuator? *Photosynthesis Research*, pages 1–16, 2020.

[3] P. Ferreira da Silva, L. Paulo, A. Barbafina, F. Elisei, F. H. Quina, and A. L. Maçanita. Photoprotection and the photophysics of acylated anthocyanins. *Chemistry–A European Journal*, 18(12):3736–3744, 2012.

[4] F. Pina, J. Oliveira, and V. de Freitas. Anthocyanins and derivatives are more than flavylium cations. *Tetrahedron*, 71(20):3107–3114, 2015.

[5] Y. Leydet, R. Gavara, V. Petrov, A. M. Diniz, A. J. Parola, J. C. Lima, and F. Pina. The effect of self-aggregation on the determination of the kinetic and thermodynamic constants of the network of chemical reactions in 3-glucoside anthocyanins. *Phytochemistry*, 83:125–135, 2012.

[6] Marco Archetti, Thomas F Döring, Snorre B Hagen, Nicole M Hughes, Simon R Leather, David W Lee, Simcha Lev-Yadun, Yiannis Manetas, Helen J Ougham, Paul G Schaberg, et al. Unravelling the evolution of autumn colours: an interdisciplinary approach. *Trends in ecology & evolution*, 24(3):166–173, 2009.

[7] A Sass-Kiss, J Kiss, P Milotay, MM Kerek, and M Toth-Markus. Differences in anthocyanin and carotenoid content of fruits and vegetables. *Food Research International*, 38(8-9):1023–1029, 2005.

[8] Emrul Kayesh, Lingfei Shangguan, Nicholas Kibet Korir, Xin Sun, Nadira Bilkish, Yanping Zhang, Jian Han, Changnian Song, Zong-Ming Cheng, and Jinggui Fang. Fruit skin color and the role of anthocyanin. *Acta physiologiae plantarum*, 35(10):2879–2890, 2013.

[9] J. A. Stavang, S. Freitag, A. Foito, S. Verrall, O. M. Heide, D. Stewart, and A. Sønsteby. Raspberry fruit quality changes during ripening and storage as assessed by colour, sensory evaluation and chemical analyses. *Scientia Horticulturae*, 195:216–225, 2015.

[10] I. Iglesias, G. Echeverria, and Y. Soria. Differences in fruit colour development, anthocyanin content, fruit quality and consumer acceptability of eight 'gala'apple strains. *Scientia Horticulturae*, 119(1):32–40, 2008.

[11] X. Zheng, C. Yue, K. Gallardo, V., J. Luby, and J. McFerson. What attributes are consumers looking for in sweet cherries? evidence from choice experiments. *Agricultural and Resource Economics Review*, 45(1):124–142, 2016.

[12] S. G. Lambert, R. E. Asenstorfer, N. M. Williamson, P. G. Iland, and G. P. Jones. Copigmentation between malvidin-3-glucoside and some wine constituents and its importance to colour expression in red wine. *Food Chemistry*, 125(1):106–115, 2011.

[13] Robert L Jackman, Rickey Y Yada, MARVIN A TUNG, and R ALEX SPEERS. Anthocyanins as food colorants—a review. *Journal of food biochemistry*, 11(3):201–247, 1987.

[14] Kin-ichi Oyama, Tomomi Yamada, Daisuke Ito, Tadao Kondo, and Kumi Yoshida. Metal complex pigment involved in the blue sepal color development of hydrangea. *Journal of agricultural and food chemistry*, 63(35):7630–7635, 2015.

[15] S. Thompson. In Vivo, In Vitro, In Vino. Presentation, CSU Department of Chemistry, 2010.

[16] Rodrigo Medel, Carezza Botto-Mahan, and Mary Kalin-Arroyo. Pollinator-mediated selection on the nectar guide phenotype in the andean monkey flower, Mimulus luteus. *Ecology*, 84(7):1721–1732, 2003.

[17] Baoqing Ding, Erin L. Patterson, Srinidhi V. Holalu, Jingjian Li, Grace A. Johnson, Lauren E. Stanley, Anna B. Greenlee, Foen Peng, H.D. Bradshaw, Michael L. Blinov, Benjamin K. Blackman, and Yao-Wu Yuan. Two MYB proteins in a self-organizing activator-inhibitor system produce spotted pigmentation patterns. *Current Biology*, 30(5):802–814.e8, 2020.

[18] T. Hoshino. Self-association of flavylium cations of anthocyanidin 3, 5-diglucosides studied by circular dichroism and 1h nmr. *Phytochemistry*, 31(2):647–653, 1992.

[19] S. González-Manzano, C. Santos-Buelga, M. Duenas, J. C. Rivas-Gonzalo, and T. Escribano-Bailon. Colour implications of self-association processes of wine anthocyanins. *European Food Research and Technology*, 226(3):483–490, 2008.

[20] A. Padayachee, G. Netzel, M. Netzel, L. Day, D. Zabaras, D. Mikkelsen, and M. J. Gidley. Binding of polyphenols to plant cell wall analogues–part 1: Anthocyanins. *Food Chemistry*, 134(1):155–161, 2012.

[21] J. Mendoza, J. Oliveira, P. Araújo, N. Basílio, N. Teixeira, N. F. Brás, and V. de Freitas. The peculiarity of malvidin 3-o-(6-op-coumaroyl) glucoside aggregation. intra and intermolecular interactions. *Dyes and Pigments*, 180:108382, 2020.

[22] E. S. Sadlowski. *pH-dependent Anthocyanin Reactions in Micellar and Copigmented Solutions*. PhD thesis, Colorado State University, 1985.

[23] Henry D Schreiber, Andrew H Jones, Corinne M Lariviere, Kelly M Mayhew, and Judith B Cain. Role of aluminum in red-to-blue color changes in Hydrangea macrophylla sepals. *Biometals*, 24(6):1005–1015, 2011.

[24] Kŏsaku Takeda, Miyuki Kariuda, and Hideaki Itoi. Blueing of sepal colour of hydrangea macrophylla. *Phytochemistry*, 24(10):2251–2254, 1985.

[25] Masaaki Shiono, Naohiro Matsugaki, and Kosaku Takeda. Structure of the blue cornflower pigment. *Nature*, 436(7052):791–791, 2005.

[26] Kumi Yoshida, Mihoko Mori, and Tadao Kondo. Blue flower color development by anthocyanins: from chemical structure to cell physiology. *Nat. Prod. Rep.*, 26:884–915, 2009.

[27] Florian P Schiestl and Steven D Johnson. Pollinator-mediated evolution of floral signals. *Trends in ecology & evolution*, 28(5):307–315, 2013.

[28] Karl von Frisch. *Der farbensinn und Formensinn der Biene.* Fischer, 1914.

[29] N Hempel de Ibarra, Misha Vorobyev, and Randolf Menzel. Mechanisms, functions and ecology of colour vision in the honeybee. *Journal of Comparative Physiology A*, 200(6):411–433, 2014.

[30] Martin J Tovée. Ultra-violet photoreceptors in the animal kingdom: their distribution and function. *Trends in ecology & evolution*, 10(11):455–460, 1995.

[31] Hansjochem Autrum et al. Die spektrale empfindlichkeit einzelner sehzellen des bienenauges. *Zeitschrift für vergleichende Physiologie*, 48(4):357–384, 1964.

[32] R Menzel. Electrophysiological evidence for different colour receptors in one ommatidium of the bee eye. *Zeitschrift für Naturforschung C*, 30(9-10):692–694, 1975.

[33] R Menzel and M Blakers. Colour receptors in the bee eye— morphology and spectral sensitivity. *Journal of Comparative Physiology A*, 108(1):11–13, 1976.

[34] R Menzel and Werner Backhaus. Colour vision in insects. *Vision and visual dysfunction*, 6:262–293, 1991.

[35] James K Bowmaker and HJk Dartnall. Visual pigments of rods and cones in a human retina. *The Journal of physiology*, 298(1):501–511, 1980.

[36] Christa Neumeyer. Chromatic adaptation in the honeybee: successive color contrast and color constancy. *Journal of Comparative Physiology*, 144(4):543–553, 1981.

[37] Annette Werner, Randolf Menzel, and Chr Wehrhahn. Color constancy in the honeybee. *Journal of Neuroscience*, 8(1):156–159, 1988.

[38] Randolf Menzel. Untersuchungen zum erlernen von spektralfarben durch die honigbiene (Apis mellifica). *Zeitschrift für vergleichende Physiologie*, 56(1):22–62, 1967.

[39] Martin Giurfa, M Vorobyev, P Kevan, and R Menzel. Detection of coloured stimuli by honeybees: minimum visual angles and receptor specific contrasts. *Journal of Comparative Physiology A*, 178(5):699–709, 1996.

[40] Martin Giurfa, Misha Vorobyev, Robert Brandt, Britta Posner, and Randolf Menzel. Discrimination of coloured stimuli by honeybees: alternative use of achromatic and chromatic signals. *Journal of Comparative Physiology A*, 180(3):235–243, 1997.

[41] Natalie Hempel de Ibarra and Misha Vorobyev. Flower patterns are adapted for detection by bees. *Journal of Comparative Physiology A*, 195(3):319–323, 2009.

[42] Silke Stach and Martin Giurfa. How honeybees generalize visual patterns to their mirror image and left–right transformation. *Animal Behaviour*, 62(5):981–991, 2001.

[43] Silke Stach and Martin Giurfa. The influence of training length on generalization of visual feature assemblies in honeybees. *Behavioural brain research*, 161(1):8–17, 2005.

[44] M. Lehrer, George Adrian Horridge, S. W. Zhang, and Raghavendra Gadagkar. Shape vision in bees: innate preference for flower-like patterns. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 347(1320):123–137, 1995.

[45] Natalie Hempel de Ibarra, Keri V Langridge, and Misha Vorobyev. More than colour attraction: behavioural functions of flower patterns. *Current Opinion in Insect Science*, 12:64–70, 2015. Neuroscience * Special Section: Insect conservation.

[46] Marinus L de Jager, Edward Willis-Jones, Samuel Critchley, and Beverley J Glover. The impact of floral spot and ring markings on pollinator foraging dynamics. *Evolutionary Ecology*, 31:193–204, 2017.

[47] Natalie Hempel de Ibarra, Martin Giurfa, and Misha Vorobyev. Detection of coloured patterns by honeybees through chromatic and achromatic cues. *Journal of Comparative Physiology A*, 187(3):215–224, 2001.

[48] Anne S. Leonard and Daniel R. Papaj. 'x' marks the spot: The possible benefits of nectar guides to bees and plants. *Functional Ecology*, 25(6):1293–1301, 2011.

[49] Eben Goodale, Edward Kim, Annika Nabors, Sara Henrichon, and James C Nieh. The innate responses of bumble bees to flower patterns: separating the nectar guide from the nectary changes bee movements and search time. *Naturwissenschaften*, 101(6):523–526, 2014.

[50] Alan Mathison Turing. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 237(641):37–72, 1952.

[51] Alfred Gierer and Hans Meinhardt. A theory of biological pattern formation. *Kybernetik*, 12(1):30–39, 1972.

[52] Hans Meinhardt and Alfred Gierer. Pattern formation by local self-activation and lateral inhibition. *BioEssays*, 22(8):753–760, 2000.

[53] Hans Meinhardt. Turing's theory of morphogenesis of 1952 and the subsequent discovery of the crucial role of local self-enhancement and long-range inhibition. *Interface Focus*, 2(4):407–416, 2012.

[54] Richard V Craster and Roberto Sassi. Spectral algorithms for reaction-diffusion equations. *arXiv preprint arXiv:1810.07431*, 2018.

[55] Henry Adams, Tegan Emerson, Michael Kirby, Rachel Neville, Chris Peterson, Patrick Shipman, Sofya Chepushtanova, Eric Hanson, Francis Motta, and Lori Ziegelmeier. Persistence

images: A stable vector representation of persistent homology. *Journal of Machine Learning Research*, 18, 2017.

[56] Barbara Di Fabio and Massimo Ferri. Comparing persistence diagrams through complex vectors. In *Image Analysis and Processing—ICIAP 2015: 18th International Conference, Genoa, Italy, September 7-11, 2015, Proceedings, Part I 18*, pages 294–305. Springer, 2015.

[57] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

[58] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20:273–297, 1995.

[59] Christopher Tralie, Nathaniel Saul, and Rann Bar-On. Ripser.py: A lean persistent homology library for python. *The Journal of Open Source Software*, 3(29):925, Sep 2018.

[60] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[61] Carlos Adrian Vargas Aguilera. extrema2.m. MATLAB Central File Exchange, 2016.

[62] John Burkardt. voronoi_mountains.m. University of South Carolina, 2019.

[63] Tom Walsh. Voronoi Cones. Geogebra. URL: https://www.geogebra.org/m/HNxm3pxf.

[64] Ehud Meron. Vegetation pattern formation: The mechanisms behind the forms. *Physics Today*, 72(11):30–36, 2019.

[65] Max Rietkerk and Johan van de Koppel. Regular pattern formation in real ecosystems. *Trends in Ecology & Evolution*, 23(3):169–175, 2008.

[66] E. Gilad, J. von Hardenberg, A. Provenzale, M. Shachak, and E. Meron. Ecosystem engineers: From pattern formation to habitat creation. *Phys. Rev. Lett.*, 93:098105, Aug 2004.