

甲醛鱼类图像分类中 K-NN 方法与 GLCM 特征提取的应用

学士学位论文

DIAH AYU LARASATI

178160018



信息工程专业

工程学院

University of Medan Area

棉兰

2021 年

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Document Accepted 18/5/26

Access From (repositori.uma.ac.id)18/5/26

甲醛鱼类图像分类中 K-NN 方法与 GLCM 特征提取的应用

学士学位论文

作为获得 University of Medan Area 工程学院学士学位条件之一提交的论文

作者:

DIAH AYU LARASATI

178160018

信息工程专业

工程学院

Universitas Medan Area

棉兰

2021 年

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Document Accepted 18/5/26

Access From (repositori.uma.ac.id)18/5/26

Judul Skripsi : Penerapan Metode K-NN dan Ekstraksi Fitur GLCM dalam
Mengklasifikasi Citra Ikan Berformalin

Nama : Diah Ayu Larasati

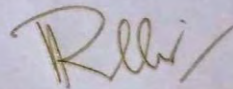
NPM : 178160018

Fakultas : Teknik

Disetujui Oleh
Komisi Pembimbing



Muhathir, ST., M.Kom
Pembimbing I



Rizki Muliono, S.Kom., M.Kom
Pembimbing II

Diketahui :



R. Rizki Muliono, S.Kom., M.Kom
Dekan Fakultas Teknik



Rizki Muliono, S.Kom., M.Kom
Ka. Bina/WDI

Tanggal Lulus : 31 Agustus 2021

HALAMAN PERNYATAAN

Saya menyatakan dengan sungguh-sungguh bahwa tugas akhir ini adalah hasil penelitian, pemikiran dan presentasi asli saya sendiri. Saya tidak mencantumkan tanpa pengakuan bahan yang telah diterbitkan atau ditulis oleh orang lain sebelumnya, atau sebagai bahan yang telah diajukan untuk gelar diploma di Universitas Medan Area atau perguruan tinggi lainnya.

Apabila dikemudian hari terdapat kejanggalan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik sesuai dengan ketentuan yang berlaku di Universitas Medan Area.

Demikian pernyataan ini saya buat.

Medan, 25/04/2022

Yang membuat pernyataan,



Diah Ayu Larasati
178160018

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR/SKRIPSI/TESIS UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademika Universitas Medan Area, saya yang bertanda tangan di bawah ini:

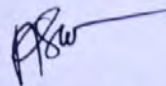
Nama : Diah Ayu Larasati
NPM : 178160018
Fakultas : Teknik
Program Studi : Informatika
Jenis Karya : Tugas Akhir

Demi pengembangan ilmu pengetahuan, setuju untuk memberikan kepada Universitas Medan **Hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul:

Penerapan Metode K-NN dan Ekstraksi Fitur GLCM Dalam Klasifikasi Citra Ikan Berformalin

Bersama dengan perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti yang bersifat *non-eksklusif* ini, Universitas Medan Area berhak menyimpan, mengalihkan media/format, mengelola dalam bentuk database, memelihara dan mempublikasikan tugas akhir/tesis/skripsi saya selama saya tetap menyebut nama saya sebagai pencipta/penulis dan sebagai pemilik hak cipta. Demikian Surat pernyataan ini saya buat dengan sebenarnya.

Dibuat di: Medan
Pada tanggal: 25 April 2022
Yang menyatakan



(Diah Ayu Larasati)

摘要

鱼类含有丰富的蛋白质，某些种类鱼类所含的蛋白质甚至高于肉类。印度尼西亚是一个约 75%国土由海洋组成的国家，因此印度尼西亚拥有巨大的海洋渔业潜力，每年渔业资源产量可达到 6500 万吨。鱼类属于容易腐败的食品，因此许多渔民和鱼类销售者使用具有危害性的化学物质——甲醛。甲醛属于致癌物质，即能够诱发癌症的物质。因此，作者开展了基于图像的甲醛鱼类分类研究，采用 K-NN 算法中的曼哈顿距离公式以及 GLCM 特征提取方法进行分类。本研究使用了两种不同类型的鱼类，即罗非鱼和 Tamban 鱼。根据研究结果，罗非鱼图像的分类准确率达到 100%，Tamban 鱼图像的分类准确率为 61%，其精确率 (Precision) 为 0.63，召回率 (Recall) 为 0.63，F1-Score 为 0.61。

关键词：罗非鱼、Tamban 鱼、甲醛、K-NN、GLCM

ABSTRACT

Fish has high protein, even certain types of fish contain higher protein than meat. Indonesia is a country where 75% of its territory consists of oceans, which causes Indonesia to have a large potential for marine fish, each year the fishery resources in Indonesia reach 65 million tons. Fish is a perishable food. This causes many fishermen and fish sellers to use the dangerous chemical formalin. Formalin is a carcinogenic substance, meaning a substance that can trigger cancer. Therefore, the author conducted a study to classify formalin fish based on its image using the K-NN algorithm with the manhattan distance formula and GLCM feature extraction. This study used two different types of fish, namely tilapia and tamban fish. And based on the research, the accuracy of the image of tilapia is 100% and the image of tamban fish is 61%, with a precision of 0.63, a recall of 0.63 and an F1-Score of 0.61.

Keyword: Mujair Fish, Tamban Fish, Formalin, k-NN and GLCM

生平简介

DIAH AYU LARASATI, 于 1999 年 8 月 19 日出生于棉兰市, 是 Rasiman 先生与 Laila Yusra 女士夫妇两名子女中的长女。

作者于 2011 年完成位于 Deli Serdang 县 Percut Sei Tuan 区 101783 Saentis 小学的小学教育。同年, 作者继续在 Deli Serdang 县 Percut Sei Tuan 区第三国立初级中学就读, 并经过三年的学习于 2014 年毕业。随后, 作者于 2014 年进入位于 Deli Serdang 县 Percut Sei Tuan 区的第一国立高级中学继续深造, 并于 2017 年毕业。

同年, 作者继续在私立高等院校深造, 即进入 Universitas Medan Area 工程学院信息学专业学习。在大学期间, 作者积极参与多项活动, 例如学生创造力计划 (PKM), 并参与了若干以期刊形式发表的研究。2020 年, 作者在位于 Deli Serdang 县 Percut Sei Tuan 区的 PAB 8 Saentis 私立高级中学完成了实习工作。

前言

谨向全能的真主安拉 (Allah SWT) 致以崇高的赞美与感谢, 感谢其赐予的恩典与慈爱, 使作者得以完成题为《甲醛鱼类图像分类中 K-NN 方法与 GLCM 特征提取的应用》的学士学位论文。

作者深知, 本论文的完成离不开各方所给予的鼓励、支持、帮助、指导、建议与合作。因此, 作者借此机会向以下人士致以最诚挚的感谢:

1. 感谢全能的真主安拉 (Allah SWT), 凭借其恩典与指引, 本论文得以顺利完成。
2. 感谢 Prof. Dr. Dadan Ramdan, M. Eng., M. Sc., 作为 Universitas Medan Area 校长。
3. 感谢 Dr. Rahmad Syah, S.Kom., M. Kom., 作为 Universitas Medan Area 工程学院院长。
4. 感谢 Susilawati, S.Kom., M. Kom., 作为 Universitas Medan Area 研究与社会服务副院长。
5. 感谢 Yudi Daeng Polewangi, ST., MT., 作为 Universitas Medan Area 人力资源发展与财务行政副院长。
6. 感谢 Rizki Muliono, S.Kom., M. Kom., 作为信息工程专业主任兼第二导师, 在论文撰写过程中给予作者指导、鼓励、支持与帮助, 使本毕业论文得以顺利完成。
7. 感谢 Muhathir, S. T., M. Kom., 作为第一导师, 为作者提供了诸多意见、批评、建议与激励, 并指导作者完成本毕业论文。

8. 感谢 Cinta Rakyat 周日市场的鱼类商贩们，感谢他们提供时间与协助，为消费者提供新鲜捕获的鱼类，同时也为本研究提供研究数据材料。
9. 感谢作者的父母，始终给予作者支持、鼓励、关怀与动力，并满足作者在毕业论文撰写期间所需的一切需求。
10. 感谢学生创造力计划（PKM）的朋友与团队成员，在四年的大学生活中给予作者支持与陪伴。
11. 以及所有协助作者完成本毕业论文而无法一一列举姓名的各方人士，谨致以诚挚的感谢。

作为普通人，作者难免存在不足，因此作者深知本毕业论文仍不够完善。故此，作者诚挚期待各方提出建设性的批评与建议，以促进未来进一步的发展与完善。

棉兰，2022 年 4 月 23 日

Diah Ayu Larasati

学号：178160018

目录 (DAFTAR ISI)

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Document Accepted 18/5/26

Access From (repositori.uma.ac.id)18/5/26

第一章 引言

1.1 背景

信息的需求不仅涵盖人与人之间，也包括人与机器之间。图像中所包含的信息可以被人类以不同的方式解读，其信息价值具有主观性，取决于每个个体的需求。

图像处理技术的发展旨在为人类生活提供便利。图像处理技术的发展得益于计算机技术领域的飞速进步，尤其体现在处理速度和存储容量方面。这些进步使得实时处理数字图像数据成为可能，并能在不降低图像质量的前提下，以更小的存储容量将图像数据保存在内存中（Sulistiyawan 等，2016）。

印度尼西亚作为海洋和农业大国，其渔业蓬勃发展，这得益于该国大部分领土由海洋组成（Wijayanti & Lukitasari, 2016）。据联合国粮农组织（FAO）统计，在鱼类产量方面，印度尼西亚仅次于中国，位居全球第二。2016年，印度尼西亚的鱼类产量达到683万吨，产值高达125.3万亿印尼盾（Mardiana等人，2020）。如今，消费者在选择食用食材时更加挑剔。选择食品时最主要的考量是安全性。作为食品，鱼类深受印尼民众喜爱（Sukamandi & Sudiadnyana, 2019）。随着印尼渔业产业的蓬勃发展，受现有条件限制，渔民捕获的鱼类无法全部运往市场。鱼类是极易腐烂的食品。因此，鱼贩和渔民都会对鱼进行保鲜处理以防止腐烂（Wijayanti, 2016）。鱼类之所以容易腐烂，是因为微生物存在于其各个部位，尤其是鳃、鱼皮和内脏中（Sholihul Huda, 2017）。

另一种常见的防腐方法是使用冰块进行保鲜。这种方法成本较高且不够实用。这导致渔民和鱼贩为防止腐烂，会采取投机取巧的做法，添加福尔马林等有害化学物质。相比冰块，福尔马林使用更方便，且能使物品保持较长时间的新鲜。然

而，即使剂量极微，使用福尔马林也极其危险（Kafiar et al., 2019）。使用福尔马林腌制鱼类必须高度警惕，若食用含有福尔马林的鱼类，可能导致头晕、呕吐、呼吸道刺激，甚至喉咙灼烧感。福尔马林对健康的其他影响还可能导致某些器官（如肝脏、大脑、肾脏和中枢神经系统）受损（Tambunan et al., 2018）。有时公众对福尔马林对健康的负面影响缺乏了解和重视，甚至吸入福尔马林也会刺激鼻腔、眼睛和喉咙。因此，我们最好选择对身体有益的健康食物。选购新鲜健康的鱼类作为蛋白质来源，在购买前需对鱼类进行格外仔细的观察。因此，笔者将该问题作为毕业论文的课题，题为“**甲醛鱼类图像分类中K-NN方法与GLCM特征提取的应用**”。该系统旨在通过K-NN（*k-最近邻*）算法以及GLCM（*灰度共现矩阵*），并实现高准确率。其中，GLCM是一种专门用于测量不同图像中形状和纹理的矩阵。GLCM被选为基于对象的图像分类分析中的关键特征，有助于提高分类准确率。KNN（*k-最近邻*）是一种非常简单但有效的分类器（Haixiang et al., 2016）。选择kNN分类方法是因为它在样本量巨大的数据集上表现出色，甚至接近无限大。kNN分类可能受多种因素影响，例如k值的选择、距离度的选择等。本研究使用的鱼类图像包括罗非鱼图像——作为一种价格相对低廉且深受印尼消费者喜爱的战略性蛋白质供应商品；以及淡水鲫鱼图像——该鱼种在马卡拉海峡水域等地区仍非常丰富，且价格经济实惠，约为3000印尼盾/公斤。这两种鱼类均具备可通过特征进行识别的实体属性。通过比较不同模式下对象的特征，可以识别出它们之间的差异。对象特征的显著差异将使该对象具有较高的区分能力。因此，基于对象特征对物体/对象进行的分类可以实现较高的准确度。

1.2 问题陈述

基于研究背景，提出的问题陈述如下：

1. 如何利用K-NN方法设计系统，通过图像识别区分新鲜生鱼与经福尔马林处理的生鱼？
2. K-NN方法在分类含福尔马林鱼类图像时的准确度如何？

1.3 问题范围

在撰写毕业论文时，笔者将研究范围限定在以下方面：

1. 基于桌面端构建的鲜鱼与福尔马林处理鱼图像分类系统，采用*K-最近邻* (*K-NN*) 方法并使用曼哈顿距离公式。
2. 所使用的鱼类数据为罗非鱼和鲫鱼图像，像素尺寸为 310×231 ，图像文件格式为JPG。
3. 该系统采用*Microsoft Visual Studio*进行开发，并使用*MySQL*作为数据库来存储新鲜生鱼和经福尔马林处理的生鱼图像数据。

1.4 研究目的

本研究的目标如下：

1. 利用KNN方法构建能够区分新鲜鱼类图像与经福尔马林处理鱼类图像的系统
2. 了解KNN方法在分类含福尔马林鱼类图像时的准确度。

1.5 研究假设

在本研究中，鱼类分类使用鱼类图像作为样本数据和 *训练数据*。鱼类图像将首先使用 *训练数据* 进行训练，在完成 *训练数据* 输入后，鱼类图像将进行分类。

1.6 研究

开展本研究旨在为广大社会带来以下益处：

1. 建立能够根据图像区分新鲜鱼类和福尔马林处理鱼类的系统。
2. 增进公众对鲜鱼与福尔马林处理鱼差异的认知。
3. 通过食用新鲜且富含蛋白质的鱼类，提高公众健康标准。

1.7 的写作结构

论文的写作结构是阐述论文写作流程的思维顺序，为了便于完成研究报告，

需要制定写作结构，以便读者能够轻松理解报告从头到尾的内容。为了

便于理解，作者将写作结构分为五个章节。以下是各章节的说明：

1. 第一章 引言

第一章为引言，包含完善研究提案的内容，包括：问题背景、研究问题、研究意义及写作结构。

2. 第二章 文献综述

本章探讨文献研究、工作原理、所用材料的描述以及构建分类系统所需的各项说明。

3. 第三章 分析与设计

本章探讨研究方法论，包括研究的时间和地点、所用材料及方法、理论框架

、数据来源、研究数据类型、所用分析方法以及研究工作流程。

4. 第四章 结果与讨论

本章包含对系统工作原理、系统测试以及本研究提出的系统分析的讨论。

5. 第五章 结论与建议

本章阐述了已开展研究的核心内容，并提出建议，指出本研究中存在的不足之处，以便未来进一步发展。

6. 参考文献

参考文献列出了本研究中使用的所有参考文献，并按特定格式编排成列表。

7. 附录

本部分包含与本研究讨论内容相关的图表。

第二章 文献综述

21. 机器 学习

机器学习是无数重要应用的基础，包括网络搜索、反垃圾邮件、语音识别、产品推荐等 (Ng, 2016)。机器学习中的数据推断更侧重于变量间关系的领域。当对大规模数据进行推断时，机器学习技术显得尤为直观。这使得机器学习广受欢迎，因为在构建模型时，推断过程可以自动完成。机器学习被比作一种“工具”，就像数学中的公式一样。具体如何应用取决于所面临的问题 (J. W. G. Putra, 2020)。

机器学习由从现有数据中学习 (训练) 的模型构成。数据可能需要进行预处理，以识别并处理缺失或虚假的元素。例如，无机晶体结构数据库 (ICSD) 目前包含 188,000 条记录，虽然已检查过技术错误，但仍可能存在人为和测量误差。若要避免机器学习算法受这些错误误导，识别并消除这些错误至关重要。公众对经同行评审的科学文献中发表的实验数据缺乏可重复性及错误传播的担忧日益加剧。在某些特定领域 (如化学领域)，已制定了最佳实践和指南来解决这一问题。

机器学习模型的训练可以是监督式、半监督式或无监督式，这取决于可用数据的类型和数量。这些算法的目标是根据一组特定的输入值，构建出能够以可接受的精度预测输出值的函数。如果可用数据集仅包含输入值，则可采用无监督学习来识别数据中的趋势、模式或聚类。当输入数据量庞大，但相应的输出值数量有限时，半监督学习可能具有价值。

监督学习是这些方法中最成熟且最强大的，被广泛应用于物理学领域的机器学习研究中，例如将化学成分映射到目标属性。无监督学习虽不常见，但可用于更普

遍的数据分析和分类，或用于在大型数据集中识别先前未知的模式（Butler et al., 2018）。

2.2 数字图像

推动图像处理技术发展的最初契机，是1921年首次成功通过数字方式发送照片。

这张记录该事件的照片从*纽约市*出发，穿越大西洋，经由海底电缆传至伦敦。

由 Harry

G. Bartholomew及其同事Maynard D. McFarlane首次开发了*巴特兰电缆图像传输系统*。该系统的优势在于大幅缩短了传输时间，即原本需要数周才能完成的照片传输，通过该系统仅需3小时即可完成。该系统的缺点在于，接收方若要获得高质量图像，仍需重新打印并传输数字数据。（Sulistiyawan 等，2016）。

数字图像是一种广泛用于直接沟通或通过互联网进行沟通的数据类型（Zebua & Ndruru, 2017）。通过图像挖掘信息直至获得*输出结果*并用于特定目的的过程，被称为数字图像处理（Rianto & Harjoko, 2017）。图像处理通常指利用计算机对二维图像进行处理（Sari et al., 2017）。数字图像主要分为两种类型：*光栅图像*（位图）和*矢量图像*（Hafiz, 2019）。

图像是对物体或事物的表现、相似物或仿制品，例如（Bee et al., 2016）（Bali, 2019）：

- a. 代表镜头前自我实体的自拍照
- b. 代表体内状况的胸部X光片
- c. BMP文件中的数据代表了所描绘的内容。

数字图像以数字文件格式存储，且只能通过计算机进行处理。若需对其他类型的

图像进行计算机处理，则必须先将其转换为数字图像（放射诊断与放射治疗技术学院，2019）。

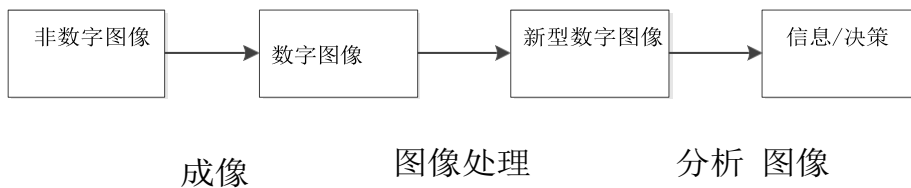


图1. 图像分析（放射诊断与放射治疗技术学院，2019年）。

2.3 鱼

鱼类蛋白质含量高，某些鱼类的蛋白质含量甚至高于肉类（Tambunan 等，2018）。从营养角度来看，优质鱼类是天然脂肪酸的来源，其中欧米伽-3脂肪酸含量最高（Febrianti & Sari, 2016）。鱼类极易发生腐败。腐败通常由鱼体内的微生物活动引起（Zamhariroh & Hanum, 2018）。一般而言，海产鱼类价格较为亲民，且在特定时期鱼类存量可能非常丰富（Kafiar et al., 2019）。本次研究重点针对罗非鱼和鲫鱼的鱼类形象进行研究。

2.3.1 罗非鱼

罗非鱼（拉丁学名：*Oreochromis Mossambicus*）能在各种水质条件下生存和繁殖，无论是高盐度水体还是低盐度水体（Rasmi et al., 2020）。与其他淡水鱼类相比，罗非鱼繁殖速度较快，这保障了罗非鱼资源的充足供应。罗非鱼的高营养价值促使人们将其作为各类食品的原料。罗非鱼每100克含79.7毫升水分、18.7克蛋白质、89千卡热量、1克脂肪、96毫克钙和1.5毫克铁（Sary & Syuhada, 2019）。

罗非鱼在市场上非常容易买到。2011年，印度尼西亚人均罗非鱼消费量达到5.18 千克 (Nurmala et al., 2020)。罗非鱼约于1936年或1939年（存在两种说法）首次在东爪哇省布利塔尔河河口被莫贾伊尔先生发现，但尚不清楚是谁引入的，因此该鱼更常被称为罗非鱼。鲤鱼和罗非鱼被列为“全球百大最危险入侵物种” (Muchlisin, 2020)。穆贾伊尔鱼是杂食性鱼类（动物和植物），如水生植物、硅藻、甲藻、绿藻、微型甲壳类和蓝藻，属于贪食性鱼类 (Pratomo et al., 2020)。

幼鱼阶段，该鱼倾向于摄食浮游生物及浮游动物 (Rasmi et al., 2020)。

罗非鱼群居生活，平静的水域、河流和淡水湖是其理想的生存环境。罗非鱼也可在水族箱中饲养，但其生长速度不会像在池塘或开放水域中养殖和生活的罗非鱼那样快 (Walingkas, Kapantow, & Ruauw 2016)，该鱼种体型扁平，呈灰色、黑色或棕色 (Dirham & Trianto, 2020)。罗非鱼最外层覆盖着鳞片，呈片状，类似于一层坚硬的皮肤，形状像质地坚硬的小片，起保护身体的作用。鱼鳞属于鳞片类，且罗非鱼属于硬骨鱼纲 (*Actinopterygii*) (Atamtajani & Amelia, 2019)。



图2. 罗非鱼

2.3.2 沙丁鱼

沙丁鱼 (*Sardinella fimbriata*) 是一种非常常见的食用鱼，分布于印度-太平洋水域 (Syah et al., 2020)。沙丁鱼含有21.77%的EPA和11.59%的DHA，含量相当高。沙丁鱼常被加工成沙丁鱼罐头、用作牲畜饲料或鱼粉原料，其富含的EPA和DHA使其在药用及保健产品领域具有开发潜力 (Nasution et al., 2019)。

沙丁鱼的鳍式如下：D. XII, C. XII, V. VI, 无胸鳍，且背鳍无分隔 (Singkam 等, 2020)。淡水鲫鱼的价格约为每公斤11,101印尼盾 (Sari et al., 2020)。淡水鲫鱼 (当地俗称 ikan tamban) 体长可达16厘米。尾部呈黑色，体侧中线呈金色 (Kurniawan et al., 2019)。



图3. 淡水鲫鱼

2.4 福尔马林

福尔马林用于杀灭细菌，因此可有效用于清洁地板、船只，以及尸体防腐处理 (Labagow 等, 2019)。蛋白质与福尔马林结合后，细菌组织会发生脱水，导致细菌细胞干燥并在表面形成新层，从而有效杀灭细菌。福尔马林形成的新层使其对细菌的侵袭具有抵抗力 (Sugiarti & Aminah., 2019)。

福尔马林被禁止用于食品 (Astuti & Tebai, 2018)。福尔马林会迅速与消化道和呼吸道的黏膜层发生反应 (Wardani & Mulasari, 2016)。体内福尔马林含量过高可能引发过敏及胃部刺激, 福尔马林具有致癌性 (致癌物) 并会改变细胞/组织功能 (致突变物)。若被摄入, 摄入者将出现呕吐、血尿, 甚至导致死亡 (Kimia, 2016) (Ali & Gustina, 2019)。一般而言, 人体可耐受的福尔马林浓度上限为0.1 mg/l (Sammulia等, 2020)。当人体免疫力低下时, 即使低浓度的福尔马林也可能对健康造成不良影响 (Tambunan等, 2018)。

福尔马林之所以被选作食品防腐剂, 是因为其价格实惠、易于获取且使用方便 (C H Yulianti, 2021)。卫生部第722/Menkes/Per/IX/88号条例以及 2012年第33号卫生部条例规定 (Fauziyya & Saputro, 2020), 禁止在食品中使用福尔马林 (Sammulia et al., 2020)。

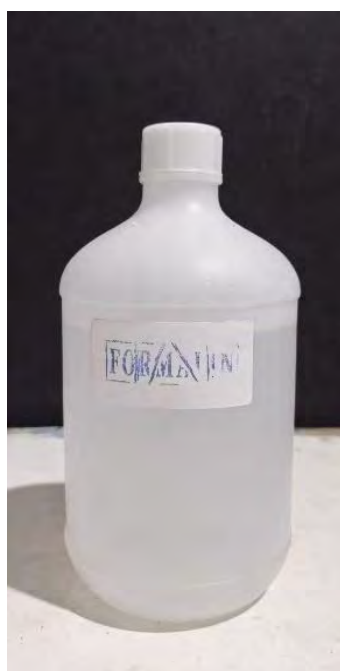


图4. 福尔马林

2.5 K-最近邻 (K-)

k-邻域法 (k-NN) 是一种数据挖掘技术, 被认为是数据挖掘领域排名前十的技术之一 (Seidl, 2009)。k-NN算法是一种被称为实例基方法的技术 (S. N. Wibowo et al., 2017) (Salvador-Meneses et al., 2019)。k-NN分类器通过将无标签观测值分配给与其最相似的有标签实例所属的类别, 来对无标签观测值进行分类 (Rohkhimet al., 2020) (Anderio & Johan, 2019)。观测值的特征被收集到训练集和测试集中。通常考虑多个邻居会很有帮助。KNN数据分为两部分: 训练数据和测试数据 (Putra, 2020)。训练数据被投影到多维空间中, 其中每个维度代表训练数据中的特征; 根据训练数据的分类结果, 该多维空间被划分为若干区域 (Kusuma et al., 2017) (Satrian & Gusrianty, 2020)。最近邻 (NN) 算法与k-最近邻 (kNN) 算法。NN仅是kNN的特例, 其中k = 1。

$$\langle \mathbf{x}^{[i]}, \mathbf{y}^{[i]} \rangle \in \mathcal{D} (|\mathcal{D}| = n) \dots\dots\dots (1)$$

正因如此, kNN也被称为懒惰学习算法。所谓懒惰学习算法, 是指将训练样本的处理推迟到进行预测时 (再次强调, 这种训练实际上只是存储训练数据 (Raschka, 2018))。随后, 为了进行预测 (类别标签或连续目标), KNN算法会寻找查询点的k个最近邻, 并根据这k个最近邻 (相似点) 计算类别标签 (分类) 或连续目标 (回归) (Laksana & Sulianta, 2017)。

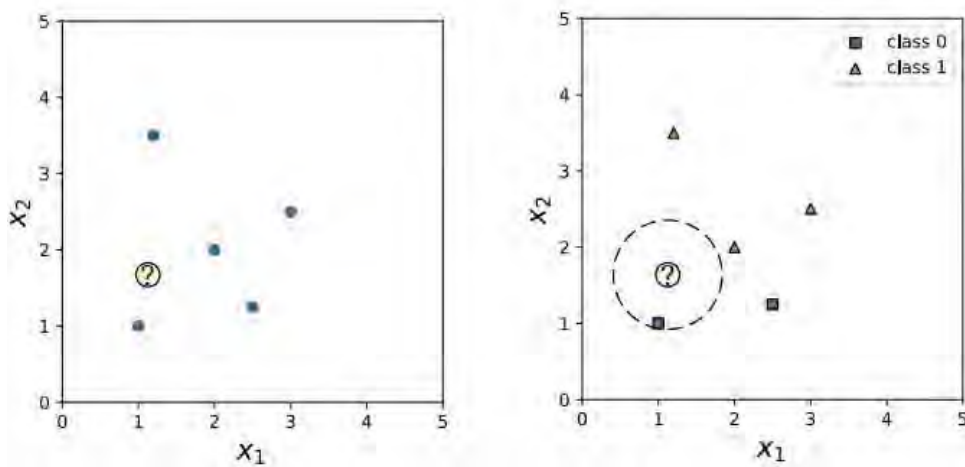


图 5. K-NN 示意图 (Laksana & Sulianta, 2017)

图5展示了二维空间（特征 x_1 和 x_2 ）中最近邻分类算法的示意图。在左侧子面板中，训练样本以蓝色点表示，而我们想要进行分类的查询点则以问号标记。在右侧子面板中，显示了类别标签，虚线表示查询点的最近邻，假设采用欧几里得距离度量。预测的类标签即为训练集中最近数据点的类标签（此处为类 0）。

kNN 算法通过将输入数据与输出相似的数据进行关联来对数据进行分类。在开发 kNN 模型时，需要调整两个参数设置：最近邻邻居的数量 k ，以及数据点之间的距离。距离可以通过欧几里得距离、曼哈顿距离或明科夫斯基距离来计算。为了训练模型，将训练集中的数据放置在与数据维度兼容的坐标系中 (Saçlı et al., 2019)。曼哈顿距离的计算公式见下文方程1 (et al., 2017)：

$$d(x,y) = \sum_{i=1}^m |x_i - y_i| \dots\dots\dots(2)$$

KNN 需要整数 k 、训练数据集以及用于衡量邻近度的度量标准。KNN 的优缺点 (Amra & Maghari, 2017) :

优点:

1. 易于理解和应用。
2. 训练速度快。

缺点:

1. 属于懒惰的学习者
2. 对局部数据结构敏感。

2.6 灰度共现矩阵 (GLCM)

M. Haralick 提出了一种名为 GLCM 的特征提取方法。GLCM 定义了灰度图像中两个相邻像素之间的关联 (Acar., 2016), GLCM 图像的纹理特征是通过统计具有特定值和特定空间关系的像素数量来计算的 (Korchiyne et al., 2014)。

图像特征提取是提取目标对象特征信息以区别于其他对象的阶段。提取后的特征可作为分类阶段中区分对象的输入参数 (Gade & Vyavahare, 2018)。纹理 GLCM 同时考虑两个像素之间的关系, 即参考像素和邻近像素 (Hall-Beyer, 2017)。通常在构建 GLCM 矩阵时会使用 4 个方向, 即角度 $\theta = 0^\circ$ 、 45° 、 90° 和 135° 。对于每个方向, 在给定距离 d 和角度 θ 的选定值下, 会生成一个 GLCM 矩阵 (Irawan et al., 2018) (Muhathir et al., 2021) (Sudibyo et al., 2018)。本研究中使用的 4 个纹理特征方程为:

a. 对比度

计算整个图像中相邻像素之间的亮度差异 (S. A. Wibowo et al., 2016)。

$$contrast = \sum_{n=1}^L \{ \sum_{|i-j|=n} GLCM(i, j) \} \quad (3)$$

b. 能量

能量是衡量矩阵中成对强度集中度的特征。计算能量的公式为 (Aferi et al., 2018) :

$$energy = \sum_{i,j} p^2(i, j) \dots \dots \dots (4)$$

c. 熵

熵是一种标量，可定义为随机性的统计度量。它对于识别输入图像的特征非常有帮助。(Usha & Perumal, 2019)

$$entropy = \sum_{i,j} P(i, j) \log(P(i, j)) * P(i, j) \quad (5)$$

d. 逆差分矩 (IDM)

IDM 用于衡量颜色均匀性的影响 (Sari 等, 2018)。

$$IDM = \sum_{i=1}^L \sum_{j=1}^L \frac{GLCM(i,j)}{1+GLCM(i-j)^2}$$

2.7 Microsoft Visual Studio

Visual Studio 的应用范围包括：基于 Windows 的应用程序、网站、Web 应用程序、Web 服务以及智能手机应用程序。Visual Studio 支持 36 种不同的编程语言。

Visual Studio 的优势在于，程序员在编译后可直接在 IDE 中查看更新，无需退出 IDE 并单独运行应用程序。Visual Studio 的另一优势是基础版“*Visual Studio Community Edition*”可免费获取，使其成为初学者的理想 IDE。Visual Studio 的缺点在于，其用户界面起初可能令人困惑，需要一定时间来“适应”。另一个缺点是，尽管微软开发团队频繁为 IDE 推送更新，但系统仍会出现崩溃，尤其是在编写复杂应用程序时 (Popescu & Costache, 2019)

2.8 MySQL

MySQL 数据库由多个表组成。表用于存储数据。*CREATE TABLE* 语句既能创建表，又能定义其结构。该语句可直接输入到“mysql>”提示符下，也可写入文件中，稍后提交给 MySQL。MySQL 支持多种数据类型。

MySQL 视图本质上是一种将 SELECT 语句封装为可重用虚拟表的方法，通过引用视图即可获取数据，而无需重复执行相关的 SELECT 语句。视图最常与 *连接操作* 结合使用 (Satoto et al., 2017)。

2.9 统一建模语言 (UML)

UML 为软件工程师提供了可视化软件密集型系统并促进思想交流的工具 (Robles et al., 2017)。

2.9.1 用例图

用例图 由 Ivar Jacobson 于 1986 年提出。*用例* 是一种有用的系统分析方法，用于识别和明确系统需求。

2.9.2 数据流图

数据流图（DFD）中有四种符号

，即：

1. 椭圆代表外部实体：指系统控制范围之外的人员或其他对象。
2. 圆形或圆角矩形：代表系统内的流程。
3. 箭头：代表数据流。可以是电子数据、实物，或两者兼有。
4. 开放式矩形：代表数据存储（Aleryani, 2016）。

2.9.3 活动图

UML 活动图用于建模控制流和数据流。它阐明了特定活动和操作的顺序，以及操作或用例中的具体行为。活动图基本上由活动和转换组成。一个活动定义了由有序单元序列所描述的行为，其基本元素是操作。最常见的操作类型包括：

调用操作、调用行为、发送、接收事件、接收调用、响应、创建、销毁以及抛出异常。每种操作都用于表示相应的行为。（Belghiat & Chaoui, 2018）。

2.9.4 类图

在面向对象的分析与设计，类图是最重要的实体。它定义了系统中存在的对象类型，并阐明了系统内部类之间的静态关系。类图可以展示类的操作和属性，以及适用于对象连接的约束（Aldaiej 等, 2018）。

2.9.5 序列图

序列图确定了顺序或优先级，即消息从一个对象流向另一个对象的顺序。序列图用于检查系统中特定功能调用的顺序（Gulia & Choudhury, 2016）。

2.10 实体关系图

实体关系图由基本元素及其衍生元素组成。基本元素包括实体、属性以及实体间的关系。实体代表系统中定义的元素，如与存储信息相关的人员、对象或事件，并与存储的信息相关。实体可以归入类。类是对具有共同属性的系统组件的结构化描述 (Eachová & Trebuňa, 2019)。

第三章

分析与设计

3.1 运行中的系统分析

传统市场中食品添加福尔马林的现象依然猖獗。出于各种原因，商贩通常会将这种违禁物质添加到商品中，而完全不顾购买者食用这些商品可能面临的风险。

易腐食品是商贩敢于在商品中掺入甲醛的主要原因，鱼类是极易腐烂的商品之一，且鱼体内常被检测出甲醛成分。通常渔民会将捕获的鱼运往市场，但受各种限制，部分鱼类无法运至市场。无法运往市场的鱼获，渔民通常会通过腌制或加工成其他产品来保存。然而，这一过程耗时较长，因此不择手段的渔民便选择在商品中添加福尔马林。经过处理的鱼获随后会被重新运往市场销售。而那些原本以新鲜状态运往市场的鱼，则会直接售予消费者。

尽管鱼类是广受欢迎的食品，但商贩的鱼货常有卖不完的情况，这些鱼随后也会通过各种方式进行防腐处理。在此过程中，鱼类被添加福马林的可能性便随之产生。鱼类经防腐处理后，将重新销往市场。

3.2 系统需求分析

针对当前运行的系统，需满足以下几点：需要一个能够准确、有效地区分新鲜鱼类和含福尔马林鱼类的系统；需要对当地政府进行关于食品中使用福尔马林危害性的培训；同时需要能够支撑所构建系统运行的软硬件；需要足够的样本以最大限度地提高系统的工作效率。

3.3 所提议系统的分析

该甲醛鱼图像分类系统基于DECOPS设计，使用*Microsoft Visual Studio 2010*和*Visual Basic*编程语言构建。系统包含3个页面，其中首页包含分类菜单和数据采集功能。在使用系统时，用户首先需要输入训练数据。训练数据包含新鲜鱼类和甲醛处理鱼类的图像数据，这些数据将根据其类别进行分组。只有在输入训练数据后，才能对图像进行分类。如果鱼类图像数据已经过训练，下一步是在分类页面进行测试。在此阶段，将测试系统对图像分类的准确性。

3.4 系统开发方法

本研究采用的开发方法是软件开发生命周期（SDLC），软件遵循一个明确的生命周期，涵盖从初始到退役的所有软件产品方面。软件开发生命周期（SDLC）是软件开发中一个成熟的框架（Faizi & Rahman, 2019）。SDLC是一种系统化的高效系统开发方法，但若没有测试，其实施便无从谈起。因为SDLC虽指导系统开发流程以提升质量，却无法协助发现系统缺陷（Jindal, 2016）。软件开发生命周期（SDLC）是构建或维护软件系统的策略。（Suryantara & Andry, 2018）。

3.5 研究阶段

研究阶段包含从问题识别到研究结果及结果评估的各项研究步骤。以下是福尔马林处理鱼类图像分类研究的步骤。

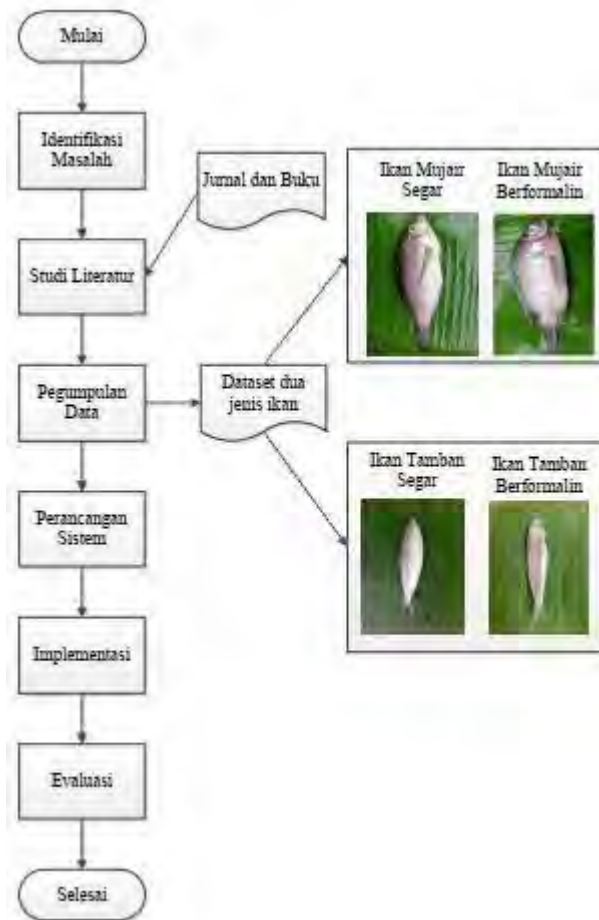


图 6. 研究阶段

上图1说明了本研究采用的各个阶段，具体说明如下：

1. 问题识别

在此阶段，根据研究者的学科领域确定研究问题，本研究的问题是鲜鱼图像与福尔马林处理鱼图像的分类。

2. 文献研究

在此阶段，作者搜寻了支持本研究课题的各类参考文献，并阅读了既往书籍及研究期刊中同类研究。

3. 数据收集

数据收集通过传统市场购买一定数量的鲜鱼进行，随后使用手机摄像头拍摄鲜鱼图像。本研究中，图像拍摄使用小米Note 9和vivo Y93手机，拍摄距离约为25 - 30厘米。完成鲜鱼图像采集后，向鲜鱼中加入适量福尔马林。在向鱼体注入福尔马林前，需先将其按50:50的比例与水混合。由于福尔马林气味刺鼻，使用前应佩戴口罩和护目镜。鱼体处理完毕后，采用与拍摄鲜鱼时相同的方式，再次对经福尔马林处理的鱼体进行图像采集。图像尺寸为310×231，文件格式为JPG，平均文件大小为60-70 KB。该系统目前尚不完善，尚无法对大尺寸图像进行分类。因此，采集的图像在使用前需先进行压缩。笔者使用*Mass Image Compressor*应用程序来压缩图像尺寸。



图7. 罗非鱼图像



图8. 鲫鱼图像

4. 系统设计

以下是鲜鱼图像和福尔马林处理鱼图像分类系统的操作流程，该流程通常以如下流程图的形式进行说明：

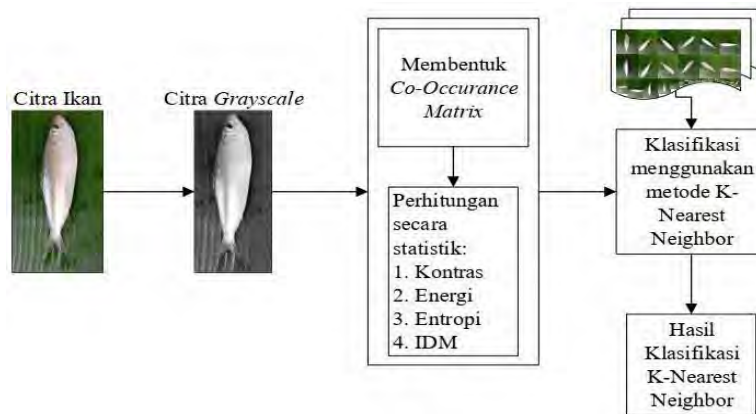


图9. 系统工作流程

本研究设计的系统基于桌面平台，可在计算机上运行。上图10展示了该系统图像分类工作流程的总体步骤，首先将原始鱼类图像数据输入系统，随后对原始鱼类图像数据进行灰度化处理，将其转换为灰度图像，此步骤旨在简化鱼类图像的像素。接下来是构建共现矩阵的阶段。GLCM的计算过程始于共现矩阵的构建阶段，即通过计算图像中的熵、对比度、能量及IDM。计算完成后，将使用K最近邻（K-NN）方法对鱼类图像进行分类（Afriyana, 2018）。

5. 实现

在此阶段，笔者采用K-NN方法和GLCM特征提取对鱼类图像进行分类。具体而言，首先计算鱼类图像的熵值、能量值、逆不同矩（IDM）及对比度，随后进行分类。

6. 评估

进行评估是为了确定所开发的系统在分类福尔马林处理过的鱼类图像时能否产生准确的结果。

3.6 KNN方法的操作步骤

下文给出了经过GLCM预处理的罗非鱼图像样本，该处理已生成其熵、对比度、能量和IDM数据，这些数据将用于KNN方法的计算过程，具体如下：

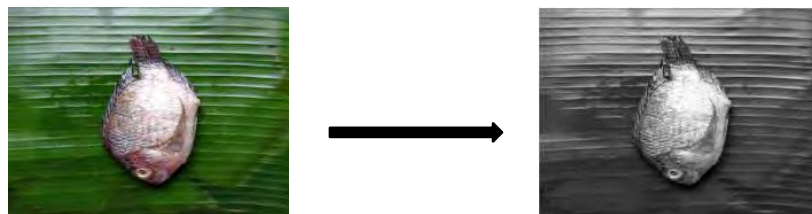


图 10. 灰度化效果

原始鱼类图像将应用灰度化处理，将其转换为灰度图像，此操作旨在简化鱼类图像的像素。

Energi	5853416
Kontras	1074356,75
Entropi	35670628689246,6
Idm	5381995,24305063

图11. GLCM数据

根据上图12，得到以下数据：

1. 能量 = 5853416
2. 对比度= 1074356. 75
3. 熵 = 35670628689246. 6
4. IDM = 5381995. 24305063

上述图像数据将使用kNN方法进行分类，并确定该鱼类图像属于哪个类别。以下是用于鱼类类别检索过程的图像训练数据。

表1. 图像数据

能量	对比度	熵	IDM	说明
5853416	1231453. 25	4118097740429	537. 583. 100. 032. 958	新鲜鱼
5853416	868302. 75	4170240071010 2. 1	547, 942, 465, 969. 732	新鲜鱼

585341 6	1023178.7 5	3272506060425 8.4	539.396.526.225. 083	含福尔马林 鱼
585341 6	803375.75	3158349203895 8.5	550,354,091,324. 263	含福尔马林 鱼
585341 6	974237.75	3158349203895	543.112.697.624. 373	新鲜鱼
585341 6	1074357	3982537846567	5381995,243	?

kNN 方法的步骤如下：

1. 首先确定要考虑的邻居数量 (k)，例如确定3个最近的邻居 (k=3)。在确定k的数值时，如果分类数量为偶数，建议使用奇数k；如果分类数量为奇数，则建议使用偶数k。

2. 接下来，计算每个邻居的距离，然后将结果按从小到大的顺序排列。如下所示。

1. 距离 1

其中所用数据如下：

1. 能量 = 5853416

2. 对比度 = 1231453.25

3. 熵 = 41180977404293

4. IDM = 5375831.00032958

使用K-NN曼哈顿距离计算距离:

$$d(x, y) = |5853416 - 5853416|$$

$$+ |1074356.75 - 1231453.25|$$

$$+ |41180977404293 - 35670628689246.6|$$

$$+ |5375831.00032958 - 5381995.24305063|$$

$$d(x, y) = |0| + |-157096.5| + |5510348715046.4| + |-6164,242721050|$$

$$d(x, y) = 0 + 157096,5 + 5510348715046,4 + 6164,242721050$$

$$d(x, y) = 5510348878307.14$$

2. 距离 2

其中使用的数据如下:

1. 能量 = 5853416

2. 对比度 = 868302.75

3. 熵 = 41702400710102.1

4. IDM = 5479424.65969732

使用K-NN曼哈顿距离计算距离:

$$d(x, y) = |5853416 - 5853416| + |868302.75 - 1231453.25|$$

$$+ |41702400710102.1 - 35670628689246.6|$$

$$+ |5479424.65969732 - 5381995.24305063|$$

$$d(x, y) = |0| + |-363150.5| + |6031772020855.5| + |97429.41664669|$$

$$d(x, y) = 0 + 363150.5 + 6031772020855.5 + 97429.41664669$$

$$d(x, y) = 6031772481435.42$$

3. 距离 3

其中使用的数据如下:

1. 能量 = 5853416
2. 对比度 = 1023178.75
3. 熵 = 32725060604258.4
4. IDM = 545393965.26225083

使用K-NN曼哈顿距离计算距离:

$$d(x, y) = |5853416 - 5853416| + |1023178.75 - 1231453.25|$$
$$+ |32725060604258.4 - 35670628689246.6|$$
$$+ |545393965.26225083 - 5381995.24305063|$$

$$d(x, y) = |0| + |-208274.5| + |-2945568084988.2| + |540011970.01919|$$

$$d(x, y) = 0 + 208274.5 + 2945568084988.2 + 540011970.01919$$

$$d(x, y) = 2946108305232.72$$

4. 距离 4

其中使用的数据如下:

1. 能量 = 5853416
2. 对比度 = 803375.75
3. 熵 = 31583492038958.5
4. IDM = 5503540.91324263

使用K-NN曼哈顿距离计算距离:

$$\begin{aligned}d(x, y) &= |5853416 - 5853416| + |803375.75 - 1231453.25| \\ &+ |31583492038958.5 - 35670628689246.6| \\ &+ |5503540.91324263 - 5381995.24305063| \\ d(x, y) &= |0| + |-428077.5| + |-4087136650288.1| + |121545.67019| \\ d(x, y) &= 0 + 428077.5 + 4087136650288.1 + 121545.67019 \\ d(x, y) &= 4087137199911.27\end{aligned}$$

5. 距离 5

其中使用的数据如下:

1. 能量 = 5853416
2. 对比度 = 974237.75
3. 熵 = 39825378465678
4. IDM = 5431126.97624373

使用K-NN曼哈顿距离计算距离:

$$\begin{aligned}d(x, y) &= |5853416 - 5853416| + |74237.75 - 1231453.25| \\ &+ |39825378465678 - 35670628689246.6| \\ &+ |5431126.97624373 - 5381995.24305063| \\ d(x, y) &= |0| + |-1157215.5| + |4154749776431.4| + |49131.7332| \\ d(x, y) &= 0 + 1157215.5 + 4154749776431.4 + 49131.7332 \\ d(x, y) &= 4154750982778.63\end{aligned}$$

3. 对数据距离值进行排序，并根据最小距离值 k 确定最近邻。

表 2. 排序值

能量	对比度	熵	IDM	曼哈顿距离	顺序	说明
585341 6	123145 3.25	1231453. 25	53758 31	5510348878 307.14	4	鲜鱼
585341 6	86830 2.75	4170240 0 710102.1	54794 25	6031772481 435.42	5	新鲜鱼
585341 6	102317 8.75	3272506 0 604258.4	53939 65	2946108305 232.72	1	含福尔马林鱼
585341 6	803375.7 5	3158349 2 038958.5	55035 41	4087137199 911.27	2	含福尔马林鱼
585341 6	974237.7 5	3567062 8689246, 60	54311 27	4154750982 778,63	3	新鲜鱼

4. 然后确定最近邻居的类别。当 $k \leq 3$ 时，将其归为“是”类别。因此第 3、4 和 5 行属于“是”类别。

表 3. 确定类别

曼哈顿距离	距离排序	是否属于3-NN	3-NN的“是”类别
5510348878307, 14	4	否 ($k > 3$)	-
6031772481435, 42	5	不 ($k > 3$)	-
2946108305232, 72	1	是 ($k < 3$)	含福尔马林鱼
4087137199911, 27	2	是 ($k < 3$)	福尔马林鱼
4154750982778, 63	3	是 ($k = 3$)	新鲜鱼

5. 在确定预测值时，采用最近邻中简单的多数类别。

表 4. 图像说明

能量	对比度	熵	IDM	说明
5853416	1231453. 2 5	41180977404293, 00	537. 583. 100. 032 . 958	新鲜鱼
5853416	868302. 75	41702400710102. 1	547, 942, 465, 969 . 732	新鲜鱼
5853416	1023178. 7 5	32725060604258. 4	539. 396. 526. 225 . 083	福尔马林鱼
5853416	803375. 75	31583492038958.	550, 354, 091, 324	含福尔马林鱼

		5	.263	
5853416	974237.75	39825378465678, 00	54, 311, 267, 624. 373	新鲜鱼
5853416	1074356.7 5	35670628689246, 60	5381995, 243	含福尔马林鱼

根据多数类别中“含福尔马林鱼” > “鲜鱼”的数量，得出新数据的预测结果为“含福尔马林鱼”。

3.7 用例图

下图是使用kNN方法进行鱼类图像分类系统的用例图。

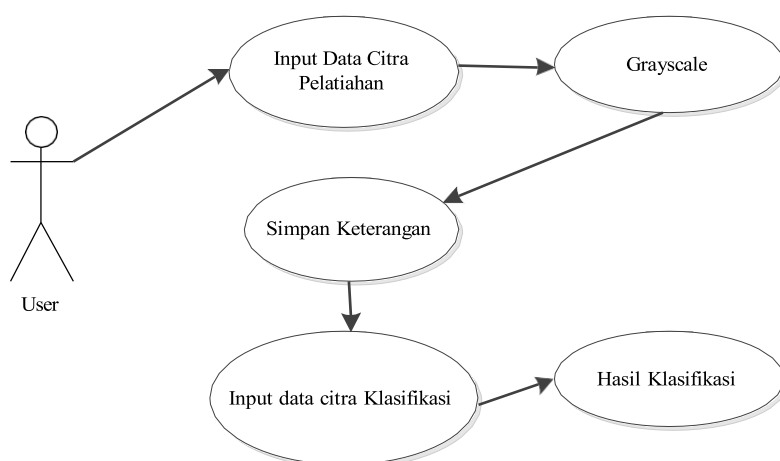


图 12. 用例图

3.8 实体关系图 (ERD)

本研究中用于构建基于MySQL系统数据库的ERD设计如下。

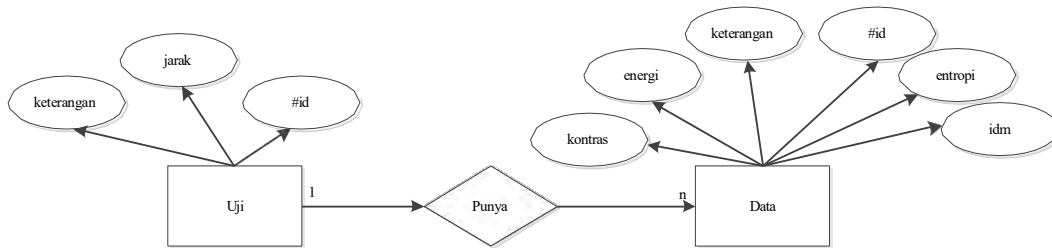


图 13. 实体关系图

3.9 表结构

在所构建的系统中设计的数据库表结构如下。

1. 数据表

数据表包含输入到系统中的所有训练数据。

表 5. 数据表

字段名称	类型	大小	说明
Id	整数	11	主键
说明	Varchar	20	非空
能源	Double	-	非空
对比	双	-	非空
熵	双精度	-	非空
IDM	双精度	-	非空

2. 测试表

测试表包含以下鱼类图像分类系统的测试数据。

表 6. 测试表

字段名称	类型	大小	说明
Id	整数	11	主键
说明	Varchar	20	非空
距离	Double	-	非空

3.10 用户界面 (UI) 设计

基于上述设计，笔者为冷冻鱼图像和鲜鱼图像分类应用程序设计了用户界面 (UI)。以下是已设计好的用户界面 (UI)。

1. 主页面

该图像分类系统的首页设有2个按钮，用户点击后将分别跳转至不同的页面。数据采集按钮将引导用户进入应用程序的数据采集或训练数据页面，而分类按钮则将引导用户进入应用程序的分类或测试数据页面。



图 14. 主页设计

2. 数据挖掘页面

在数据挖掘部分，用户将在同一页面上找到许多依次排列的功能。首先，用户可以通过“浏览图像”按钮导入想要进行训练的图像。接下来，使用“灰度”按钮将图像转换为灰度图像。之后，用户可以点击“处理 GLCM按钮，计算所输入图像的GLCM结果。在GLCM计算过程中，左侧的空白框中将显示特定角度 (θ) 的中间结果，即 0° 、 45° 、 90° 、 135° 。最终最终计算结果将显示在“处理GLCM”按钮下方。接下来，用户可以为已导入的图像添加说明，并保存GLCM计算结果。

中间角度结果 (θ)

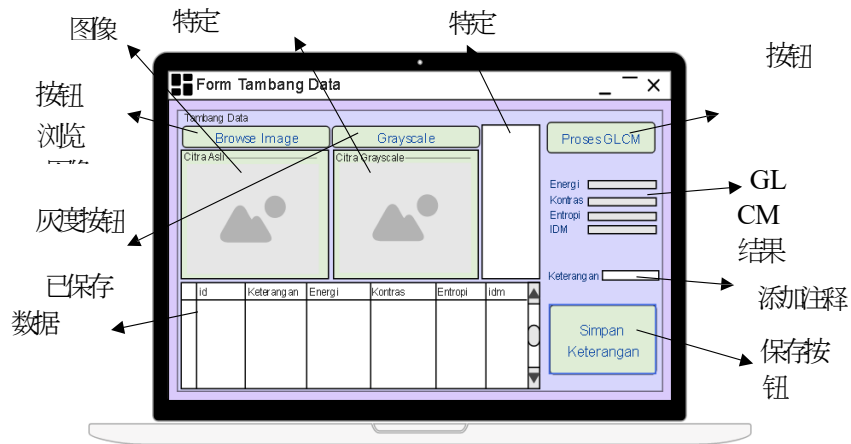


图 15. 数据挖掘页面

3. 分类页面

在分类页面上有几个按钮，其功能如下。

1. 浏览图片 : 插入图片
2. 分类 : 分类过程
3. 关闭 : 退出页面



图 16. 分类页面

在关闭页面之前，应用程序会先向用户请求确认。



图 17. 退出确认

第五章 结论与建议

5.1 结论

根据已进行的研究结果，可以得出以下结论：

1. 本研究设计了一套基于桌面端的福尔马林处理鱼类图像分类系统，采用 Visual Basic 编程语言，在 Visual Studio 环境下开发，并使用 MySQL 作为系统数据库。
2. 通过采用 K 最近邻 (K-NN) 方法和 灰度共现矩阵 (GLCM) 特征提取技术，对两种鱼类的鲜鱼与福尔马林处理鱼图像进行分类，结果显示：罗非鱼的分类准确率为 61%，罗非鱼的分类准确率为 100%。根据 $F1$ -Score 计算结果可知，采用 K -最近邻法对鱼类图像进行分类，其结果相当理想，与朴素贝叶斯法的 $F1$ -Score 对比为 0.7:0.61:1.00。在罗非鱼图像分类中，40 个测试数据均实现了完美分类。本研究认为，kNN 方法在分类中表现更优且更稳定，尤其在罗非鱼图像分类方面。当然，从错误率仍较高这一角度来看，本研究结果尚未达到最佳。这可能是由于数据采样过程尚不完善，将在后续研究中予以改进

5.2 建议

该福尔马林处理鱼类图像分类系统仍存在诸多不足，为促进系统后续发展，笔者提出以下建议：

1. 进一步优化系统，因为目前该系统尚无法对大尺寸图像进行分类。本研究使用的图像数据大小不超过 100 Kb。在后续系统开发中，希望系统能够对更大尺寸、更高质量的图像进行分类。

2. 本次研究仅使用了2种鱼类作为数据。在后续开发中，可使用更多种类和更丰富的鱼类。
3. 目前，系统在图像测试中仍需通过“选择照片/图像数据”功能逐一选择照片。在后续开发中，希望系统能够在第一轮测试过程中对多张图像进行分类。

参考文献列表

- Acar, E., (2016). Extraction Of Texture Features From Local Iris Areas By GLCM And Iris Recognition System Based On KNN. *European Journal of Technic EJT*. 6(1), 44–52.
- Aferi, F. A., Purboyo. T. W., & Saputra, R. E., (2018). Cotton texture segmentation based on image texture analysis using gray level run length and euclidean distance. *International Journal of Applied Engineering Research*, 13(1), 6915–6923.
- Afriyana Y., R. Purnamasari & Patmasari, R. (2018). Spine Curvature Disorders Based on Digital Medical Image With Gray Level Co-Occurrence Matrix (GLCM) and K-Nearest Neighbor (KNN). *e-Proceeding of Engineering*. 34675–4682,.
- Aldaej, R., Alfowzan, L., Alhashem, R., Alsmadi, M. K., Al-Marashdeh, I., Badawi, U. A., Alshabanah, M., Alrajhi, D., & Tayfour, M. (2018). Analyzing, Designing and Implementing a Web-Based Auction online System. *International Journal of Applied Engineering Research*, 13(10), 8005–8013. <http://www.ripublication.com>
- Aleryani, A. Y. (2016). Comparative Study between Data Flow Diagram and Use Case Diagram. *International Journal of Scientific and Research Publications*, 6(3), 124–127.
- Ali, H., & Gustina, M. (2019). Analisis Kandungan Zat Pengawet Pada Jajanan Bakso Di Sekolah Dasar Wilayah Kecamatan Ratu Agung Kota Bengkulu. *Journal of Nursing and Public Health*, 7(1), 59–63. <https://doi.org/10.37676/jnph.v7i1.788>
- Amra, I. A. A., & Maghari A. Y.A. (2017). Students Performance Prediction Using KNN and Naïve Bayesian. *Internasional Cpnference on Information Technology (ICIT)*. 909–913.
- Anderio, J., & Johan. (2019). Sistem Pendukung Keputusan Pemilihan Material Bangunan Berdasarkan Kesesuaian Budget Konsumen Menggunakan K-Nearest Neighbor (KNN). (Studi Kasus : Toko Bangunan AJJ). *Jurnal Mahasiswa Aplikasi Teknologi Komputer Dan Informasi*, 1(1). 12-19.
- Astuti, I., & Tebai, P. (2018). Analisis Formalin Ikan Teri (*Stolephorus sp*) Asin Di Pasar Tradisional Kabupaten Gorontalo. *Gorontalo Fisheries Journal*, 1(1), 43. <https://doi.org/10.32662/.v1i1.105>
- Akademi Teknik Radiodiagnostik dan Radioterapi. (2019). Pengolahan Citra Digital. Atro Bali.
- Atamtajani, A. S. M., & Amelina, D. R. (2019). Eksplorasi Limbah Sisik Ikan Mujair Sebagai Perhiasan. *Jurnal ATRAT* ,7(1), 21–32.
- Bali, B. (2019). Modul Pengolahan C Itra Digital. *akademi Teknik Radiodiagnostik dan Radioterapi*.
- Bee, D., Weku, W. Ch. D., & Rindengan, A., J. (2016). Aplikasi Penentuan Tingkat Kesegaran Ikan Selar Berbasis Citra Digital Dengan Metode Kuadrat Terkecil. *JdC*, 5(2), 121-130. <https://doi.org/10.35799/dc.5.2.2016.14985>
- Belghiat, A., & Chaoui, A. (2018). A graph transformation of activity diagrams into pi-verification purpose. *International Conference on Advanced of*

Software Engineering (ICAASE), 1(2), 107–114.

- Dirham & Trianto, M. (2020). Analisis Isi Lambung Ikan Mujair (*Oreochromus mossambicus*) Danau Talaga Kabupaten Donggala. *BIO-EDU : Jurnal Pendidikan Biologi*. 5(3), 118–128. <https://doi.org/10.32938/jbe>.
- Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O., & Walsh, A. (2018). Machine learning for molecular and materials science. *Nature*, 559(7715), 547–555. <https://doi.org/10.1038/s41586-018-0337-2>
- Febrianti, D. R., & Sari, R. M., (2016). Analisis Kualitatif Formalin Pada Ikan Tongkol Yang Dijual Di Pasar Lama Banjarmasin. *Jurnal Pharmascience*, 03(02), 64–68.
- Faizi, S. M., & Rahman, S. S. M. (2019). Choosing the best-fit lifecycle framework while addressing functionality and security issues. *Proceedings of 34th International Conference on Computers and Their Applications, CATA 2019*, 58, 107–116. <https://doi.org/10.29007/cfm3>
- Fauziyya, R., & Saputro, A. H. (2020). Analisis Formalin Secara Kualitatif pada Bakso dan Mie Basah di Kecamatan Sukarame, Wayhalim, dan Sukabumi. *KOVALEN: Jurnal Riset Kimia*, 6(3), 218–223. <https://doi.org/10.22487/kovalen.2020.v6.i3.15333>
- Gade, A. A., & Vyavahare, A. J. (2018). Feature Extraction using GLCM for Dietary Assessment Application. *International Journal of Multimedia and Image Processing (IJMIP)*, 8(2), 409–413. <https://doi.org/10.20533/ijmip.2042.4647.2018.0050>
- Gulia, S., & Choudhury, T. (2016). An efficient automated design to generate UML diagram from Natural Language Specifications. *Proceedings of the 2016 6th International Conference - Cloud System and Big Data Engineering, Confluence 2016*, 641–648. <https://doi.org/10.1109/CONFLUENCE.2016.7508197>
- Hafiz, A. (2019). Steganografi Berbasis Citra Digital Untuk Menyembunyikan Data Menggunakan Metode Least Significant Bit (Lsb). *Jurnal Cendikia*, 17(1), 194–198.
- Haixiang, G., Yijing, L., Yanan, L., Xiao, L., & Jinling, L. (2016). BPSO-Adaboost-KNN ensemble learning algorithm for multi-class imbalanced data classification. *Engineering Applications of Artificial Intelligence*, 49, 176–193. <https://doi.org/10.1016/j.engappai.2015.09.011>
- Hall-Beyer, M. (2017). GLCM Texture: A Tutorial v. 3.0. *Arts Research & Publications*, 2017–03, 75. <https://prism.ucalgary.ca/handle/1880/51900> <http://hdl.handle.net/1880/51900>
- Hastuti, E. (2017). Gambaran Penggunaan Formalin Pada Ikan Asin Dipasar Karangayu Kota Semarang. *Jurnal Keperawatan dan Kesehatan Masyarakat (STIKES)* 2(5).
- Irawan, C., Ardyastiti, E. N., Setiadi, D. R. I. M., Rachmawanto, E. H., & Sari, C. A. (2018). A survey: Effect of the number of GLCM features on classification accuracy of lasem batik images using K-nearest neighbor. *2018 International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2018*, 33–38. <https://doi.org/10.1109/ISRITI.2018.8864443>

- Jindal, T. (2016). Importance of Testing in SDLC. *International Journal of Engineering and Applied Computer Science*, 01(02), 54–56. <https://doi.org/10.24032/ijeacs/0102/05>
- Kafiar, F. P., Salim, I., & Djarwo, C. F. (2019). Identifikasi Kandungan Formalin Pada Ikan Segar Bernilai Ekonomis Tinggi Yang Terdapat di Pasar Tradisional Kota Jayapura. *Prosiding Seminar Hasil Penelitian Pengembangan Ipteks dan Seni*, 53(9), 1689–1699.
- Korchiyne, R., Farssi, S. M., Sbihi, A., Touahni, R., & Tahiri Alaoui, M. (2014). A Combined Method of Fractal and GLCM Features for MRI and CT Scan Images Classification. *Signal & Image Processing : An International Journal*, 5(4), 85–97. <https://doi.org/10.5121/sipij.2014.5409>
- Kusuma, S. F., Pawening, R. E., & Dijaya, R. (2017). Otomatisasi klasifikasi kematangan buah mengkudu berdasarkan warna dan tekstur. *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, 3(1), 17. <https://doi.org/10.26594/r.v3i1.576>
- Labagow, E., Warouw, F., & Akili, R. H., (2019). Kandungan Formalin Pada Tahu Putih Di Pasar Pinasungkulan Karombasan Kota Manado 2019. *Kesmas*, 8(7), 67–71.
- Lachová, K., & Trebuňa, P. (2019). Modelling of electronic kanban system by using of entity relationship diagrams. *Acta Logistica-International Scientific Journal About Logistics*, 6(3), 63–66. <https://doi.org/10.22306/al.v6i3.115>
- Laksana, E. A., & Sulianta, F. (2017). Analisis dan studi komparatif algoritma klasifikasi genre musik. *Seminar Nasional Teknologi Informasi Dan Multimedia 2017*, 2(1), 67–72.
- Mardiana, R., Lidyawati, & Zulfikri, M. (2020). Identifikasi Formalin Pada Ikan Segar di Pelabuhan Pendaratan Ikan Idi Rayeuk Kabupaten Aceh Timur. *Journal of Pharmaceutical and Health Research*, 1(3), 77–82. <http://ejournal.seminar-id.com/index.php/jharma/article/view/597>
- Muchlisin, Z. A. (2020). Analisis Kebijakan Introduksi Spesies Ikan Asing Di Perairan Umum Daratan Provinsi Aceh. *Jurnal Kebijakan Sosial Ekonomi Kelautan Dan Perikanan*, 1(1), 79. <https://doi.org/10.15578/jksekp.v1i1.9256>
- Nasution, E., Sudayarti, E., & Ardiani, F. (2019). Prevention of stunting in schoolchildren through the utilization of local food in the form of tamban fish (*Spratelloides gracilis*) and spinach at Rugemuk Village Labu beach district. *ABDIMAS TALENTA: Jurnal Pengabdian Kepada Masyarakat*, 4(2), 706–711. <https://doi.org/10.32734/abdimestalenta.v4i2.4215>
- Ng, A. (2018). Machine learning Yearning. *Deeplearning.ai*. https://doi.org/10.1007/978-981-10-1509-0_9
- Nurmala, A. P., Santoso, H., & Syauqi, A. (2020). Uji Organoleptik Ikan Mujair (*Oreochromis Mossambicus*) yang Direndam dengan Kulit Manggis (*Garcinia Mangostana* L) sebagai Pengawet Alami. *Jurnal Ilmiah Sains Alami*, 3(1), 1–9.
- Pandey, A., & Jain, A. (2017). Comparative Analysis of KNN Algorithm using Various Normalization Techniques. *International Journal of Computer Network and Information Security*, 9(11), 36–42. <https://doi.org/10.5815/ijcnis.2017.11.04>

- Popescu, T., & Costache, T. (2019). Designing a Windows Program for Controlling Dc-Motors Using Microsoft Visual Studio and Arduino Ide. *Journal of Industrial Design and Engineering Graphics, Tomo 14, N*, 29–34.
- Pratomo, G. N., Nurcahyo, H., & Firdaus, N. R. (2020). profil Fermentasi Ikan Mujair (*Oreochromis mossambicus*) DENGAN PENAMBAHAN NaCl. *Al-Kauniah: Journal Biologi. 13(2)*, 158–166.
- Putra, J. W. G. (2020). Pengenalan Konsep Pembelajaran Mesin dan Deep Learning. *Computational Linguistics and Natural Language Processing Laboratory, 4*, 1–235. <https://www.researchgate.net/publication/323700644>
- Putra, R. C. (2020). Pembangunan Perangkat Pendeteksi Jenis Gerakan Raket Bulu Tangkis Dengan Algoritma KNN dan SVM. *Teknika, 9(2)*, 113–120. <https://doi.org/10.34148/teknika.v9i2.291>
- Rohkhim, A. N., Al-ghifari, M. H., Rinjani L. O. M. Z., & Damastuti, N., (2020). Algoritma K-NN Dalam Mengklasifikasikan Daerah Potensi Pemain Bulutangkis Tunggal Putra Di Indonesia. *Jurnal Insand Comtech. 6(2)*, 1-7.
- Raschka, S. (2018). STAT 479: Machine Learning Lecture Notes. *Department of Statistics*. https://sebastianraschka.com/pdf/lecturenotes/stat479fs18/02_knn_notes.pdf
- Rasmi, K. A., Dewi, A. A. Y., & Artini, N. P. R. (2020). Identifikasi Parasit Pada Ikan Mujair (*Oreochromis ossambicus*) di Daerah Taman Pancing Denpasar Selatan. *Bali International Scientific Forum (BISF), 1(1)*, 12–17.
- Rianto, P., & Harjoko, A. (2017). Penentuan Kematangan Buah Salak Pondoh Di Pohon Berbasis Pengolahan Citra Digital. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems), 11(2)*, 143. <https://doi.org/10.22146/ijccs.17416>
- Robles, G., Ho-Quang, T., Hebig, R., Chaudron, M. R. V., & Fernandez, M. A. (2017). An extensive dataset of UML models in GitHub. *IEEE International Working Conference on Mining Software Repositories, May*, 519–522. <https://doi.org/10.1109/MSR.2017.48>
- Sammulia, S. F., Poluan, T., & Yusri, Y. F., (2020). Analisis Kualitatif Kandungan Formalin Pada Tahu di Pasar Jodoh Kota Batam. *Jurnal Endurance, 5(1)*, 144. <https://doi.org/10.22216/jen.v5i1.4585>
- Saçlı, B., Aydınalp, C., Cansız, G., Joof, S., Yilmaz, T., Çayören, M., Önal, B., & Akduman, I. (2019). Microwave dielectric property based classification of renal calculi: Application of a kNN algorithm. *Computers in Biology and Medicine, 112*(January). <https://doi.org/10.1016/j.compbiomed.2019.103366>
- Kurniawan., Asmarita., & Supratman, O. (2017). Identifikasi Jenis Ikan (Penamaan Lokal, Nasional Dan Ilmiah) Hasil Tangkapan Utama (Htu) Nelayan Dan Klasifikasi Alat Penangkap Ikan Di Pulau Bangka Provinsi Kepulauan Bangka Belitung. *Akuatik: Jurnal Sumberdaya Perairan, 13(1)*, 42–51. <https://doi.org/10.33019/akuatik.v13i1.1107>
- Salvador-Meneses, J., Ruiz-Chavez, Z., & Garcia-Rodriguez, J. (2019). Compressed kNN: K-nearest neighbors with data compression. *Entropy, 21(3)*, 1–20.

<https://doi.org/10.3390/e21030234>

- Sari, F. Y., Pranoto, Y. S., & Purwasih, R. (2020). Analisis Usaha Ikan Asin (Studi Kasus Desa Rebo Kecamatan Sungailiat Kabupaten Bangka). *Journal of Integrated Agribusiness*, 2(1), 20–36. <https://doi.org/10.33019/jia.v2i1.1489>
- Sari, J. I., Sulindawati, & Sitihang, H. T. (2017). Implementasi Penyembunyian Pesan Pada Citra Digital Dengan Menggabungkan Algoritma Hill Cipher Dan Metode Least Significant Bit (LSB). *Jurnal Mantik Penusa*, 1(2), 1–8. <http://ejournal.pelitanusantara.ac.id/index.php/mantik/article/view/253>
- Sari, Y., Alkaff, M., & Pramunendar, R. A. (2018). Classification of coastal and Inland Batik using GLCM and Canberra Distance. *AIP Conference Proceedings*, 1977(June). <https://doi.org/10.1063/1.5042901>
- Sary, R., & Syuhada, A. (2019). Study of Fish Drying Process Using Multilevel Shelves with Wood Fuel. *Prosiding SNTTM*.18.1-5
- Satoto, K. I., Isnanto, R. R., Kridalukmana, R., & Martono, K. T. (2017). Optimizing MySQL database system on information systems research, publications and community service. *Proceedings - 2016 3rd International Conference on Information Technology, Computer, and Electrical Engineering, ICITACEE 2016*, 1–5. <https://doi.org/10.1109/ICITACEE.2016.7892476>
- Satrian, B. & Gusrianty (2020). Penerapan Algoritma K-Nn untuk Klasifikasi Gamers Usia Sekolah. *Jurnal Mahasiswa Aplikasi Teknologi Komputer Dan Informasi*, 2(1), 19–23.
- Seidl, T. (2009). Nearest Neighbor Classification. *Encyclopedia of Database Systems*, 1, 1885–1890. https://doi.org/10.1007/978-0-387-39940-9_561
- Singkam, A. R., Yani, A. P., & Fajri, A. (2020). Keragaman Ikan Laut Dangkal Provinsi Bengkulu. *Jurnal Enggano*, 5(3), 1689–1699.
- Sudibyo, U., Kusumaningrum, D. P., Rachmawanto, E. H., & Sari, C. A. (2018). Optimasi Algoritma Learning Vector Quantization (Lsq) Dalam Pengklasifikasian Citra Daging Sapi Dan Daging Babi Berbasis Glcm Dan Hsv. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 9(1), 1–10. <https://doi.org/10.24176/simet.v9i1.1943>
- Sugiarti, M., & Aminah, S. (2019). Pengaruh Waktu Perendaman Air Garam Terhadap Penurunan Kadar Formalin Pada Cumi- Cumi Asin. *Jurnal Analis Kesehatan*, 8(1), 58–62.
- Sukamandi, N. M. M., & Sudiadnyana, I. W. (2019). Perbedaan Kandungan Formalin Pada Jenis Ikan Asin Peda dan Gabus Di Pasar Tradisional Kota Denpasar Tahun 2019. *Jurnal Kesehatan Lingkungan*, 9(2), 92–97.
- Sulistiyanti, S. R., Setyawan, F. A., & Komarudin, M. (2016). Pengolahan Citra Dasar Dan Contoh Penerapannya, *Teknosain*.
- Suryantara, I. G. N., & Andry, J. F. (2018). Development of Medical Record With Extreme Programming SDLC. *International Journal of New Media Technology*, 5(1), 47–53. <https://doi.org/10.31937/ijnmt.v5i1.706>

- Syah, M. S., Apriandi, A., & Putri, R. M. S., (2020). Pemanfaatan Air Limbah Rebusan Ikan Tamban (*Sardinella sp.*) Sebagai Flavor Pasta Alami. *Marinade*, 03(01), 22-33, <http://ojs.umrah.ac.id/index.php/marinade>.
- Tambunan, S. B., Sebayang, N. S., & Amin, N. (2018). Karakteristik Warna Ikan Asin Sepat Sebagai Indikator Pengawet Formalin Di Pasar Tradisional Desa Tunas Jaya Muaradua. *BIOTIK: Jurnal Ilmiah Biologi Teknologi Dan Kependidikan*, 5(2), 88. <https://doi.org/10.22373/biotik.v5i2.3015>
- Muhathir, Santoso, M. H., & Larasati, D. A., (2021). Wayang Iamge Classification Using SVM Method and GLCM Feature Extraction. *JITE (Journal of Informatics and Telecommunication Engineering)*, 4(2), 343-382.
- Usha, R., & Perumal, K. (2019). SVM classification of brain images from MRI scans using morphological transformation and GLCM texture features. *International Journal of Computational Systems Engineering*, 5(1), 18. <https://doi.org/10.1504/ijcsyse.2019.098415>
- Walingkas, I. K. Ch., Kapantow, G. H. M., & Ruauw, E. (2016). Pemanfaatan Teknologi Informasi Dan Komunikasi Terhadap Jaring Ikan Di Desa Eris, Kecamatan Eris, Kabupaten Minahasa. *Agri-sosioekonomi*, 12(2), 1-14.
- Wardani, R. I., & Mulasari, S. A., (2016). Identifikasi Formalin pada Ikan Asin yang Dijual di Kawasan Pantai Teluk Penyus Kabupaten Cilacap, *Kesmas*. 10(1), h 16.
- Wibowo, S. A., Hidayat, B., & Sunarya, U. (2016). Simulasi dan Analisis Pengenalan Citra Daging Sapi dan Daging Babi dengan Metode GLCM dan KNN. *Seminar Nasional Inovasi Dan Aplikasi Teknologi Di Industri (Seniati) 2016*, 338–343.
- Wibowo, S. N., Hidayat, D. I. B., & Arif, D. I. J. (2017). IDENTIFIKASI JENIS BATUAN BEKU MELIHAT BENTUK POLA BATUAN MENGGUNAKAN METODE DISCRETE WAVELET TRANSFORM (DWT) DAN K-NEAREST NEIGHBOR (KNN), *e-proceeding of Engineering*,. 4(2), 1677–1684.
- Wijayanti, N. S. & Lukitasari, M., (2016). Analisis pengawetan makanan dan uji organoleptik ikan asin yang beredar di pasar besar madiun. *Jurnal Florea Volume*, 3(1), 59–64.
- Yulianti, C H. (2021). Perbandingan Uji Deteksi Formalin pada Makanan Menggunakan Pereaksi Antilin dan Rapid Tes Kit Formalin (Labstest). *Journal of Pharmacy and Science*, 6(1), 53–58. <http://ejournal.akfarsurabaya.ac.id/index.php/jps/article/view/205>
- Zamhariroh, & Hanum, R. G. (2018). Efektivitas Temu Kunci (*Boesenbergia rotunda*) Terhadap Penurunan Kadar Formalin Pada Ikan Tuna (*Thunnus sp.*). *Journal of Medical Laboratory Science/ Technology*, 1(2), 68–76. <http://creativecommons.org/licenses/by/4.0/>
- Zebua, T., & Ndruru, E. (2017). Pengamanan Citra Digital Berdasarkan Modifikasi Algoritma RC4. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 4(4), 275– 282. <https://doi.org/10.25126/jtiik.201744474>

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Document Accepted 18/5/26

Access From (repositori.uma.ac.id)18/5/26