

# IoT Based Detection of Molded Bread and Expiry Prediction using Machine Learning Techniques

Muhammad Shoaib Akhtar, Tao Feng\*

School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China

## Abstract

Expiration of a bread is a very popular issue in food logistics. Due to various conditions fungal bread can cause food poisoning for consumers. As a result, nausea, diarrhea and different medical issues appear in people. For this purpose, an intelligent system required for the detection of present condition of bread is required which will help the stores and consumers. In this study, we have developed a prototype made up of Arduino Nano as a microcontroller, MQ series sensors for CO and CO<sub>2</sub> detection in shopper bags of bread in order to collect data. This data is further processed in different machine learning algorithms for the detection of current condition of bread in these stores. The data collected from these sensors was imbalanced. Data collected from sensors is then balanced by using SMOTE and Tomek Links (data balancing techniques). Furthermore, data preprocessing and feature engineering has been applied on IoT Based dataset to improve its efficiency. We have applied linear learning models for the prediction of current condition of bread. Within linear models, Gaussian Naïve Bayes has scored highest accuracy of 81.54%.

**Keywords:** IoT, ML, SVM

Received on 21 January 2022, accepted on 24 April 2022, published on 27 April 2022

Copyright © 2022 Muhammad Shoaib Akhtar *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [Creative Commons Attribution license](#), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eai.27-4-2022.173972

\*Corresponding author. Email: fengt@lut.cn

## 1. Introduction

All of the methods utilised in this thesis, including those for detecting yeast and automating the detection system, are outlined here. In addition, the research objective and methodology are laid out in this paper. This section concludes with an explanation of a thesis' structure.

There are countless physical, electrical, and electromagnetic devices all around us, all of which are used to address thousands of difficulties that arise on a daily basis. In today's commercial, medical, health, financial, education, household, and industry eras, the Internet plays a critical function. The Internet of Things is a remarkable invention since it combines both applications and the Internet (IoT). Director of MIT's Auto-ID-Centre, Kevin Ashton coined the term "IOT" back in 1999 [1].

User convenience is the primary goal of the Internet of Things (IoT). With the Internet of Things (IOT), the world is becoming smarter and more interconnected [2], [3]. Wherever it's utilized, from emergency clinics to grocery stores to security regions to banks to workplaces and labs,

the food area to instructive organizations and even homes, it makes the world a more brilliant and more savvy place. Bakery items are the primary focus of IOT in restaurants and bakeries [4], [5]. The most pressing issue is whether or not the bread being used is safe for consumption and whether or not it has expired. Carbohydrate-dense foods such as bread are popular among the general populace. Wheat bread is one of many new breads being produced to pique consumer interest [6]. Quality wheat bread must be produced by bakers who are popular with the general population. Sales will rise dramatically if the bread's quality is significantly improved. These baked goods may be contaminated with fungus, which can harm human health. Fungus may be invisible, but IOT can help if we can detect gases emitted from bread, such as carbon dioxide and methane. Gas sensors, such as MQ135 and MQ7, are used in this study to keep tabs on the quality of bread and determine if it has gone bad. Bread is a carb thick food that is every now and again ate by individuals from the general population [7], [8]. A number of breads have been developed to pique the interest of consumers, with wheat bread being one among them. Wheat bread of high quality

must be created by makers that are well-liked by the general public. A major increase in the overall quality of the bread will have a big effect on sales. Fungi feed on the organic matter that they are growing on, such as bread, in order to survive. The fungus's method of reproduction is shown by the colonies of spores that you see on bread [9], [10]. Molds such as *Aspergillus* and *Penicillium*, as well as *Mucor* and *Rhizopus*, can be discovered growing on bread products. An infected spore must first identify bread and then implant its "hyphae" into the bread's surface in order to survive in a dark, chilly environment with insufficient air circulation. Microbial mycelium, also known as a mould colony, is formed as a result of the fast spread of mould spores in the environment. It is vital to standardise, rationalise, and link the organoleptic assessment approach in order to ensure its effectiveness. Food industry and other agricultural product organoleptic tests are widely used to determine the quality of the products under consideration [11], [12].

Food poisoning symptoms such as diarrhoea, nausea, and vomiting can occur as a result of consumers' ignorance of the presence of these fungal illnesses on breads in bakeries and retail establishments. More than 200 people became ill after eating bread contaminated with the fungus *Mucor circinelloides*, which caused symptoms such as nausea, vomiting, and diarrhoea [13], [14]. Fungus is not only detrimental to human health, but it also poses a significant threat to the management of the food supply chain. A fungal infestation can have an impact on both the quality and the quantity of food produced. Consumers discard infected fungal grains because they are unappealing to them since they are infected. Fungus is responsible for between 5 and 10% of all food losses each year, depending on the region [15], [16].

As fungus on bread can be damaging to human health in addition to causing food loss, a method that can determine whether bread is safe to eat was required to determine whether bread is safe to consume. This research assisted in determining the amount of various gases present in bread and, on the basis of those quantities, determining whether the bread was expired or not [17], [18].

Organoleptic tests are those that are carried out by humans using menas of ogran as a test subject. For example, evaluating the bread by smelling it, touching it, and tasting it are all acceptable methods of testing. It is a routine test that can be dangerous to the individual who is being tested. A smart system that can detect and forecast the condition of the bread is required in order to avoid this type of situation in the future.

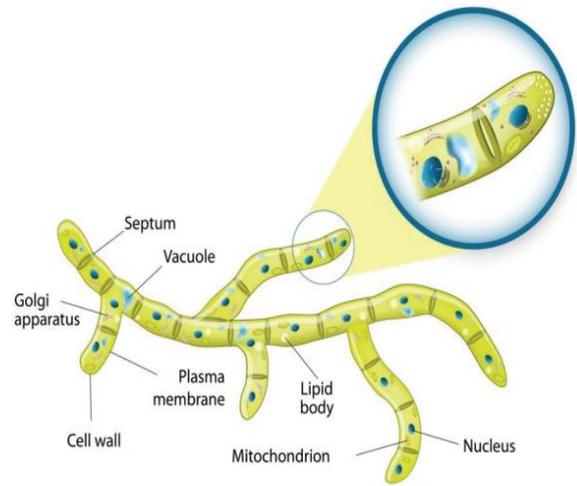


Figure 1. Fungus in Yeast.

*Aspergillus*, *Candida*, *Aspergillus*, *Histoplasmosis*, *Coccidioidomycosis*, and *Cryptococcus*'s are all examples of pathogenic fungus that can cause significant diseases in humans. *Aspergillus Niger*, *Histoplasma*, and *Cryptococcus*, *Athlete's foot*, *ringworm*, and other fungal infections can be caused by dermatophytic fungi. There are several fungi that might produce life-threatening allergic reactions if consumed in large enough quantities [19], [20]. Mycotoxins, which are harmful substances released by mold and mushrooms, are responsible for a slew of ailments. Immune-compromised people are more susceptible to the effects of these pollutant. Ergotism is caused by consumption of ergot alkaloids, which are found in mushrooms such as *Amenita*, which create fatal amatoxins. One of the most dangerous metabolites produced by *Aspergillus* is aflatoxins. Toxins enter the bloodstream, hinder protein synthesis, and weaken the body's defenses. Smut disease in maize is caused by a fungus called *Ustilago maydis*, which is a pathogen. Respiratory allergies and asthma are linked to the prevalence of fungi in the air, according to several research. Another set of disorders that can result from exposure to fungal spores include toxic pneumonitis and hypersensitivity pneumonitis, tremors, chronic exhaustion, kidney failure, and cancer. Bread molds are a group of molds that grow on bread. When mold spores reach the bread's surface, they begin to grow. Organic molecules found in bread and other meals are eaten by mold, a fungus. Three frequent bread molds are *Penicillium*, *Cladosporium*, and black bread mold [21].



**Figure 2. Fungus on Breads.**

For testing, a widespread belief is that bread can be tasted, smelled, and touched. It's a standard test that can be detrimental to the individual taking it. An intelligent system is required to identify and anticipate the status of bread in order to avoid such things. In order to develop an IoT based system that can sense the environmental variables, a huge amount of data is required. Imbalancing in a sensed data is common problem. Fungus detection systems that can be controlled remotely are designed to protect the customers using bakery products (Xu et al., 2019; Zhang et al., 2018). When exposed to fungus for an extended period of time, people can suffer from major health issues. This framework is just as valuable as anyone else's. Developing an IoT-based fungal detection system is also aimed at making it portable. When deployed in an affected area and connected to the internet, the prototype of a fungus detection system may be used to monitor fungus concentrations in real time. Following are the objectives of this study:

- a) To create a dataset of real-time carbon dioxide values from a box containing fresh bread.
- b) To create a hardware based IoT prototype containing CO<sub>2</sub> sensor and Arduino microcontroller in order to gather dataset of bread in real-time
- c) To apply machine learning algorithms in order to classify the healthy condition on the basis of supervised labelling on dataset.
- d) To evaluate machine learning models on the basis of Accuracy, Precision, Recall and F1 Score.

Previously no other study has been done on time series forecasting about healthy condition of bread based on fungal infections. It will help local general stores, bakeries in logistics and customers.

## 2. Materials and Methods

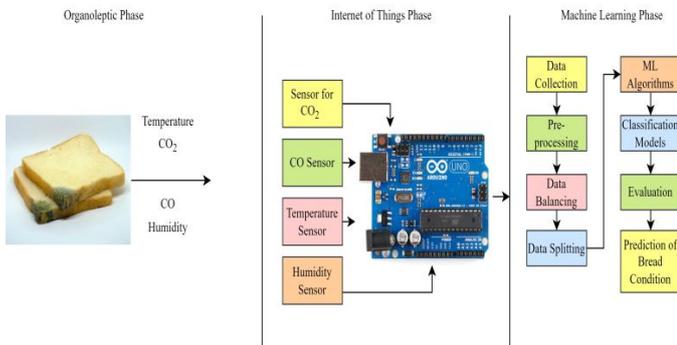
This section details the research process utilized to complete this project. Model creation (using machine learning algorithms) begins with pre-processing raw data (adjusting the dataset and handling outliers). The Internet of Things is all about making life easier for people (IoT). Smarter and more networked, the Internet of Things (IOT) is changing the world. A superior and shrewder world is made conceivable by its far-reaching use in pretty much every

area of society, from emergency clinics to grocery stores, to security regions to banks and working environments, to instructive organizations and, surprisingly, our own homes. In restaurants and bakeries, IOT is most commonly used to track and monitor baked goods. The most important question is whether or not the bread is safe to eat and whether or not it has expired. Bread and other carbohydrate-heavy foods are popular with the general population. The production of wheat bread is just one of several new breads being made in an effort to attract more customers. Wheat bread of the highest quality can only be made by well-known bakers. If the quality of the bread is greatly enhanced, sales will soar. There is a possibility that the fungus in these baked items is harmful to one's health. If we can detect gases emitted from bread, such as carbon dioxide and methane, IOT can assist in the detection of fungus. Monitoring the quality of bread and determining if it has gone bad is done by using gas sensors, such as MQ135 and M7. Carbohydrate-dense foods such as bread are popular among the general populace. Wheat bread is one of a number of breads that have been produced to capture the curiosity of consumers. Quality wheat bread must be produced by bakers who are popular with the general population. Sales will rise dramatically if the bread's quality is significantly improved. In order to exist, fungi must eat the organic materials they are growing on, such as bread. The colonies of spores you see on bread illustrate the fungus' reproductive mechanism. Bread goods can be found with moulds including *Aspergillus*, *Penicillium*, and *Mucor* and *Rhizopus*. In order to survive in a dark, chilly, and poorly ventilated environment, an infected spore must first locate bread and then insert its "hyphae" into the bread's surface. Mould colonies, also known as microbial mycelium, arise when mould spores travel quickly via the air. For the organoleptic assessment approach to be effective, it must be standardised, rationalised, and linked together. Organoleptic tests are commonly used in the food business and other agricultural industries to determine the quality of the items being considered. Consumers' ignorance of the existence of these fungal infections on breads in bakeries and retail places can cause food poisoning symptoms such as diarrhea, nausea, and vomiting. Food poisoning was reported in more than 200 people after eating bread tainted with the fungus *Mucor circinelloides*. As well as harming human health, fungi represent a serious hazard to the food supply chain itself. When it comes to food production, both the quality and the quantity might be affected by a fungus infestation. Infected fungal grains are unappetizing to consumers because they are infected. Annual food losses due to fungus range anywhere from 5 to 10 percent, depending on the region in question. Because fungus on bread can harm human health as well as cause food loss, a method for determining whether or not bread is safe to ingest was necessary. There are many gases that can be measured to determine whether a loaf of bread has gone bad or not. The results of this study helped in this regard.

## 3. Methodology

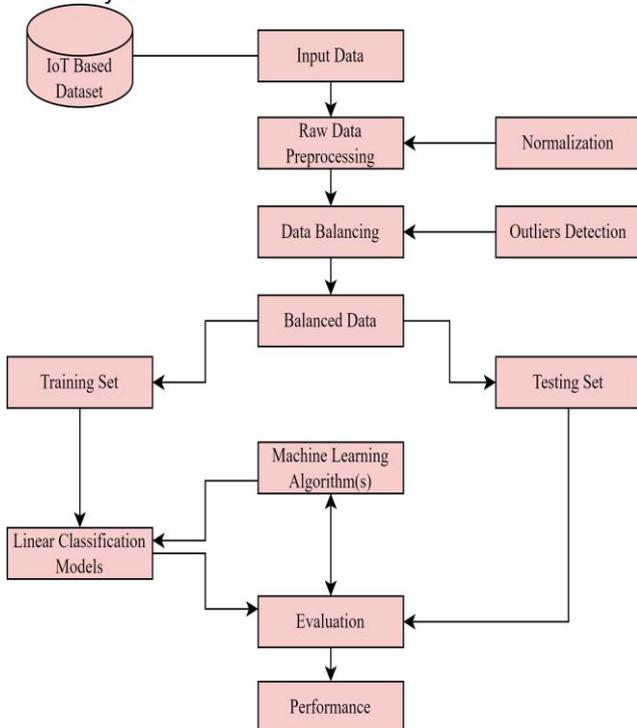
The proposed system is composed of three layers:

- (i) Sensors
- (ii) Detection Unit
- (iii) Machine Learning



**Figure 3.** Proposed Framework

There has been no single research that has brought to the low-cost detection of rotten breads. Amounts of Carbon Dioxide in the air where bread is present have been measured using sensors MQ 7 and passed to an Arduino board for analysis. The important information from sensor data is extracted using a variety of classification methods for the goal of detecting threats. Classifier algorithms such as KNN and Logistic Regression are among the most useful and successful in the field. Prior to the classification method, the classifier may employ a feature selection and outlier's detection strategy in order to obtain a more accurate answer from the dataset. The type of classifier to be used is determined by the data and the desired outcomes to be achieved. An outline of the proposed work is depicted in Figure 1 as a block diagram, which represents the summary of ideas.



**Figure 4.** Proposed Flow Work

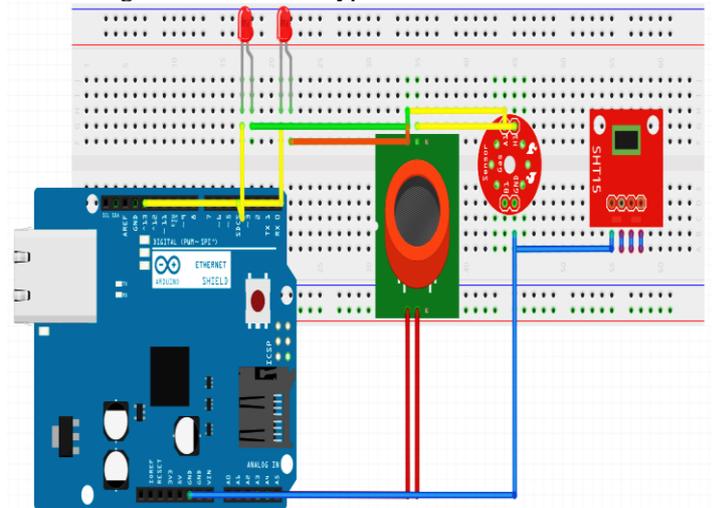
An IoT-based dataset has been used to generate the CSV input file. This data has been cleaned, balanced, and handled for outliers using data preprocessing techniques. The results have been validated using cross-validation. Machine learning models have been tested and deployed extensively. The most accurate classification models have been introduced.

### 3.1. Implementation of Hardware

Vacuum-packed breads are studied in this research utilizing our electronic technology, which is explained herein. A minimal expense and low-power utilization arrangement has been proposed by us. As an initial step, the primary module incorporates a model for information gathering, which is trailed by a subsequent that handles the UI and information capacity and handling. The data gathered tells us about the bread's current state, amongst other details. Figure depicts the two modules of our system's block diagram.

### 3.2. IoT Based Prototype

IoT Device has been presented in this study containing Arduino UNO, CO2 Sensor, CO Sensor, Humidity and Precipitation sensor in order to check the air quality under plastic bags. Figure below shows the 3D Model of IoT Based Fungal Detection Prototype:

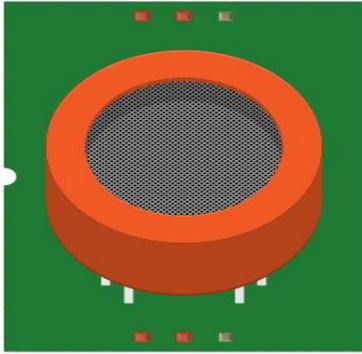


**Figure 5.** IoT Based Prototype

### 3.3. Components Details

#### 3.3.1. Carbon Dioxide Sensor and its specifications

The MQ-7 Gas Sensors are appropriate for NH3, NOx, Alcohol, Benzene, Smoke, and CO2 detection and measurement in air quality control systems.



**Figure 6.** Carbon Dioxide Sensor

Indoor and outside air hold back somewhere in the range of 400ppm and 2,000ppm CO<sub>2</sub> by volume, contingent upon the season. For applications going from assessing indoor air quality to estimating air CO<sub>2</sub>, a 0-1 percent (0-10,000ppm) CO<sub>2</sub> sensor is the most savvy and solid choice. A 5 percent CO<sub>2</sub> sensor is utilised in applications such as CO<sub>2</sub> alarms in restaurants, breweries, and areas around stored CO<sub>2</sub> tanks, among other things. Alarms must be present at 1.5 percent and 3 percent CO<sub>2</sub>, according to OSHA regulations. Other applications, such as changed atmosphere packaging, bioreactors, cryogenics, and SCUBA, necessitate the use of sensors that range from 5 to 100 percent CO<sub>2</sub>. In general, you should first determine the greatest level of CO<sub>2</sub> that you need to measure, and then select the sensor that can measure that level accurately.

**3.3.2. Carbon Monoxide Sensor with its specification**

Gas concentrations can be sensed with a simple voltage divider network. The Gas sensor is powered by 5V DC and consumes 800mW of power. As a result, it is able to measure concentrations ranging from 200 to 10000ppm for a wide range of substances.

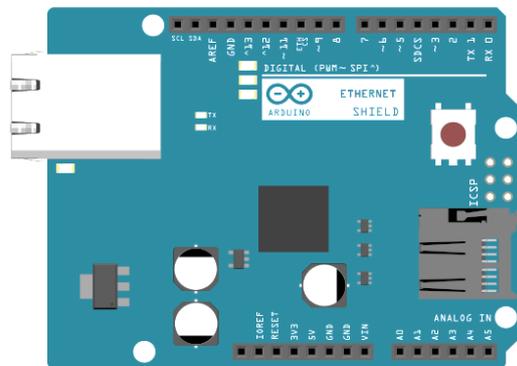


**Figure 7.** Gas Sensor for CO

This is a Carbon Monoxide Sensor made using an Arduino. Carbon Monoxide (CO) concentrations can be measured

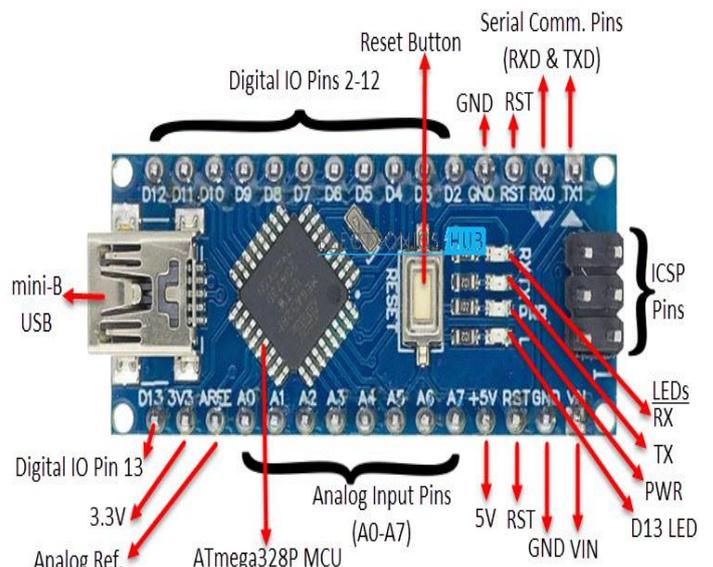
from 20 to 2000ppm using the MQ7 probe. The potentiometer can be used to modify the sensitivity. As the gas density increases, so does the generator's output. Using the potentiometer, you may change the sensitivity. As the density of the gas increases, the output increases as well; CO is measured in parts per million (ppm). As a comparison, the natural atmosphere has a concentration of 0.1 ppm. In most houses, the level is between 0.5 and 5 ppm. 5-15 ppm is a good starting point. As little as 667ppm can produce up to 50% of the effects. Four wires are required to connect the sensor. Power is a factor in two of them. The sensor's +5V terminal is connected to the Arduino board's 5V port. The sensor's GND terminal is connected to the GND terminal of the Arduino. This provides the sensor with power. The sensor's analogue and digital outputs are connected to the remaining two ports.

**3.3.3. Arduino Nano & UNO with its specifications**



**Figure 8.** Microcontroller Arduino UNO

Connecting the Arduino to a computer via USB, the Arduino environment may be accessed (IDE). After writing the code in the IDE, the user uploads it to the microcontroller, which runs the code and interacts with inputs and outputs such as sensors, motors, and lights.

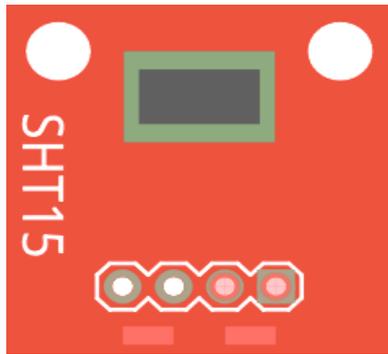


**Figure 9.** Microcontroller Arduino Nano

TX (transmit) and RX (receive) are the two labels on your board (receive). The baud rate of the board determines the speed of the flashing. During the receiving process, RX flashes. The microprocessor on each Arduino board is unique (11). Assume it is the brain of the board. Depending on the board, the Arduino's primary IC (integrated circuit) may be slightly different. The ATMEL microcontrollers are commonly used. Before uploading a new programme from the Arduino IDE, you must know what IC your board is equipped with. The IC has this information printed on the top of it. There are six analogue input pins on the Arduino UNO board, from A0 to A5. Using these pins, the microprocessor can read the analogue signal from a humidity or precipitation sensor and convert it to a digital value.

### 3.3.4. Humidity Sensor with its specifications

In a surface-mountable LCC packaging, the SHT11 is a standard version of the relative humidity and temperature sensor IC. This sensor has a small footprint, yet it has all of the necessary sensor components and signal processing to produce a digital output that is perfectly calibrated.



**Figure 10.** Sensor SHT 11x Series for Humidity

It is a digital-output, relative humidity, and temperature sensor known as the DHT-22 (also known as AM2302). Data is sent on the data pin by the capacitive humidity sensor and a thermistor that measure the temperature of the surrounding air. Pins details are as under:

3-5V

Maximum current is 2.5mA.

With a two- to five-percent degree of precision, humidity ranges from 0% to 100%

Temperatures between 40 and 80°C, with a 0.5°C precision.

There are four pins on the DHT22 sensor board (VCC, DATA, NC, and GND) that are used to interface with the sensor.

Vcc and GND are both connected to the Arduino board's GND and DATA pins, respectively.

Between the Vcc and the DATA pins, we require a 10k ohm resistor (pull-up resistor).

## 3.4. Dataset

This information has been assembled from the IoT Device. There are different (autonomous) factors and one (subordinate) variable of the dataset (Outcome).

| Humidity | Co2  | Precipitation | Co  | Expired_Bread |
|----------|------|---------------|-----|---------------|
| 75       | 582  | 20            | 1.9 | 1             |
| 55       | 7861 | 38            | 1.1 | 1             |
| 65       | 146  | 20            | 1.3 | 1             |
| 50       | 111  | 20            | 1.9 | 1             |
| 65       | 160  | 20            | 2.7 | 1             |
| 90       | 47   | 40            | 2.1 | 1             |
| 75       | 246  | 15            | 1.2 | 1             |
| 60       | 315  | 60            | 1.1 | 1             |
| 65       | 157  | 65            | 1.5 | 1             |
| 80       | 123  | 35            | 9.4 | 1             |
| 75       | 81   | 38            | 4   | 1             |
| 62       | 231  | 25            | 0.9 | 1             |
| 45       | 981  | 30            | 1.1 | 1             |
| 50       | 168  | 38            | 1.1 | 1             |
| 49       | 80   | 30            | 1   | 0             |
| 82       | 379  | 50            | 1.3 | 1             |
| 87       | 149  | 38            | 0.9 | 1             |
| 45       | 582  | 14            | 0.8 | 1             |
| 70       | 125  | 25            | 1   | 1             |

**Figure 11.** Training Set

### 3.4.1. Dataset Descriptions

Table 1. Dataset information from Bread

| Attributes            | Sensors | Description  |
|-----------------------|---------|--|
| Carbon Monoxide       | MQ 7    | The air quality sensor distinguishes alkali, nitrogen oxide, smoke, CO2, and other perilous poisons. The air quality sensor includes a little potentiometer that works with the alteration of the heap obstruction of the sensor circuit. The 5V power supply is utilized for air quality sensor.    |
| Carbon Dioxide        | MQ 135  | The air quality sensor distinguishes alkali, nitrogen oxide, smoke, CO2, and other dangerous contaminations. The air quality sensor includes a little potentiometer that works with the change of the heap opposition of the sensor circuit. The 5V power supply is utilized for air quality sensor. |
| Humidity              | DHT 11  | Data is sent out on the data pin using capacitive humidity sensors and a thermistor to measure the surrounding air (no analogue input pins needed).  |
| Outcome (Classes 0/1) | N/A     | 1 Class: Bread with Fungus<br>0 Class: Bread without Fungus  |

### 3.4.2. Dataset Attribute Statistics

Figure shows the histogram of each quality given in the dataset. It exhibits insights of dataset characteristics like CO2, CO, Outcomes and Humidity.

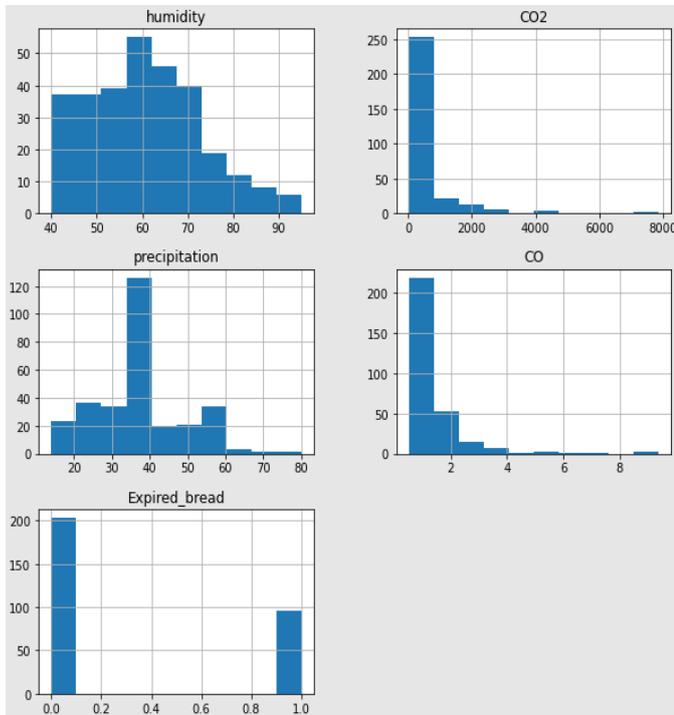


Figure 12. Histogram of each attribute

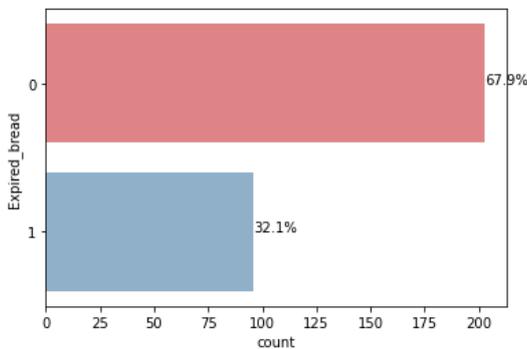


Figure 13. Relation between count and outcomes.

Figure above shows the complete count for each objective (0 or 1). The 0 implies that the bread is usable, and 1 method bread is terminated.

### 3.5. Data Pre-processing

Data Pre-processing is a significant stage in the information mining process since it can connect with the changing or dropping of information before it is utilized. Planning (cleaning and organizing) crude information with the goal that it could be utilized for making and preparing AI models is known as information pre-handling in AI. Pre-handling is fundamental to guaranteeing top notch information. Four periods of information pre-handling are utilized to make the interaction simpler: information purging, information

coordination, and information decrease. Our model's capacity to learn relies upon the quality and usable data that can be extricated from our information; in this way, pre-handling our information prior to taking care of it into a machine it is basic to learn model.

### 3.6. Features Engineering

Machine learning and statistical modelling both use include designing to pick and control pertinent attributes from crude information prior to being taken care of into an expectation model. There can never be an excess of accentuation on highlight designing with regards to ML and information science overall.

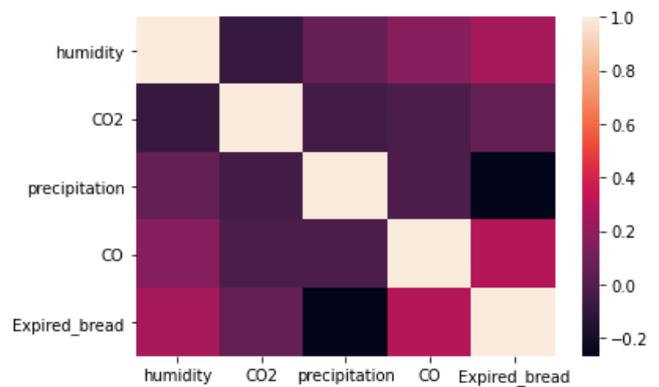


Figure 14. Correlation Matrices IoT based dataset

The fundamental goal of feature engineering is to improve algorithm performance. Feature engineering in machine learning is more than just picking and modifying features. The impact of demographic characteristics on stroke is depicted in the diagram below:

Furthermore, feature engineering improves machine learning model performance by ensuring that the dataset is compatible with the algorithm. Figure below shows the features correlation matrix.

### 3.7. Feature Scaling

Using a technique called as feature scaling, features can be normalized (Normalization). In data processing, data normalisation is also known as data preparation and is usually done during this step. For machine learning purposes, all features must be scaled so that no feature is abnormally large (centring) and all features are the same size (scaling). Because they use distances or similarities (e.g. in the form of scalar product) between data samples, K-NN and SVM are sensitive to feature modifications. It's particularly helpful while addressing a mind boggling set of conditions, like least squares, while adjusting blunders could have a critical impact. It's as yet really smart to rescale/normalize your information assuming you're utilizing Fisher LDA or Naive Bayes, as well as Decision trees and Tree-based troupe techniques (RF, XGB) that are invariant to include scaling. However, XGBoost also has a

second strategy based on linear boosting. In that instance, expanding the operation's size will be advantageous.

### 3.8. Classification Algorithms

#### 3.8.1. Support Vector Machine (SVM)

Support vector machines (SVMs), a collection of supervised learning algorithms, can be used to discover and classify outliers. There are various advantages of utilizing support vector machines. Valuable in high-layered conditions However, it can in any case be utilized in circumstances where there are a greater number of aspects than tests.

#### 3.8.2. Logistic Regression

The most appropriate regression strategy to use when the dependent variable is dichotomous (binary) is logistic regression. Measurements and numerical displaying are utilized to portray information and to make sense of the connection between a solitary ward parallel variable and at least one autonomous factors that are either ostensible, ordinal, stretch, or proportion level.

#### 3.8.3. K-Nearest Neighbors

The k-nearest neighbours (KNN) algorithm is a simple and quick to develop supervised machine learning algorithm. It can be used to solve classification and regression problems, and it can be used to solve both types of problems.

#### 3.8.4. Naïve Bayes

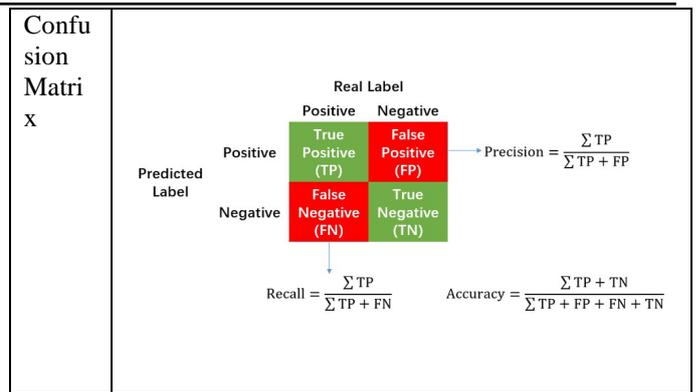
Naive Bayes classifiers are those that are based on Bayes' Theorem. It's actually a family of algorithms that all follow the same basic principle: each pair of features to be categorised is independent of the others. To get started, let's look at a dataset. Naive Bayes, a probabilistic method, is a common classification algorithm. Despite its simplicity and intuitiveness, Naive Bayes performs amazingly effectively. Spam filters in the Email app, for example, use Naive Bayes.

### 3.9. Performance Parameters

Accuracy, Precision, Recall, and F1 Score are some of the criteria that have been used to evaluate techniques. The classified and misclassified clauses have been presented in the confusion matrix. The metrics used in this study are listed in the table below:

Table 2. Description of Metrics

| Metric   | Description                             |
|----------|---|
| Accuracy | $Accuracy = \frac{TP}{(TP + TN)} * 100$ |



## 4. Results

This section demonstrates the outcomes of executing the machine learning models on the dataset after proposing the technique to forecast the bread condition. In an IoT-based dataset, Gaussian Naive Bayes is found to be a better classifier for prediction. Our findings are shown below, along with a side-by-side comparison of those of the default dataset and the balanced dataset.

### 4.1. Machine Learning Models

IoT data gathering is converted to CSV files for ML purposes. If the bread is fine, it hasn't gone bad yet. It is divided into two categories. The hotness content of carbon dioxide is increased by a steady, the carbon coefficient, isolated by the oxidized portion, and afterward duplicated by the atomic weight proportion of carbon dioxide to carbon to show up at the discharges. Sensors and organoleptic tests are used to determine the main CO2 concentration. The primary purpose of the machine learning model was to verify the accuracy of organoleptic tests based on sensor data. In datasets, the following observations were found:

Table 3. Normal and Abnormal Ranges as detected from sensors and organoleptic

| Tests               | Normal Range    | Out of Range      | Organoleptic Outcome (Condition of Bread)              |
|---------------------|-----------------|-------------------|--|
| CO2 Range           | 500 – 2500 ppm  | >2500ppm, <500ppm | Normal (500 – 2500 ppm)<br>Expired (>2500ppm, <500ppm) |
| CO Range            | 0.02 - 0.42 ppm | <0.02, >0.42ppm   | Normal (0.02 - 0.42 ppm)<br>Expired (<0.02, >0.42ppm)  |
| Humidity Range      | 30% to 50%      | <30%, >50%        | Normal (30% to 50%)<br>Expired (<30%, >50%)            |
| Precipitation Range | 70 to 80        | <70, >80          | Normal (70 to 80)<br>Expired (<70, >80)                |

#### 4.1.1 KNN

KNN Algorithm tested on bread dataset to check the outcome. It has obtained 69.23% accuracy to confidently predict the outcome of bread.

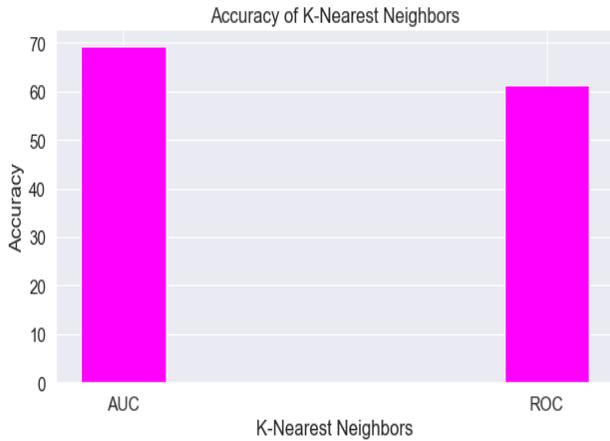


Figure 15. K-Nearest Neighbour Performance

#### 4.1.2 Logistic Regression

Logistic Regression tested on bread dataset to check the outcome. It has obtained 72.31% accuracy to confidently predict the outcome of bread.

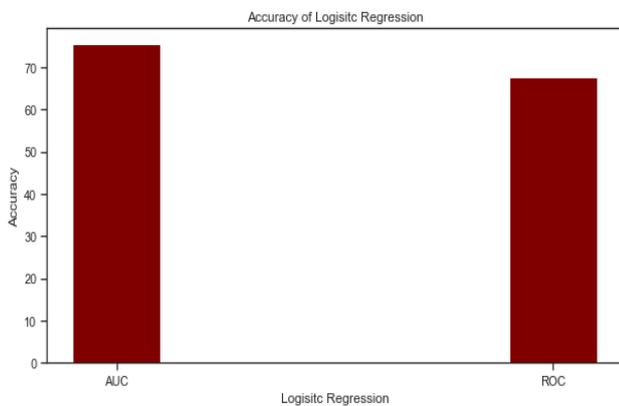


Figure 26. Performance of Logistic Regression

#### 4.1.3 Naïve Bayes

Naive Bayes tested on bread dataset to check the outcome. It has obtained 81.54% accuracy to confidently predict the outcome of bread.

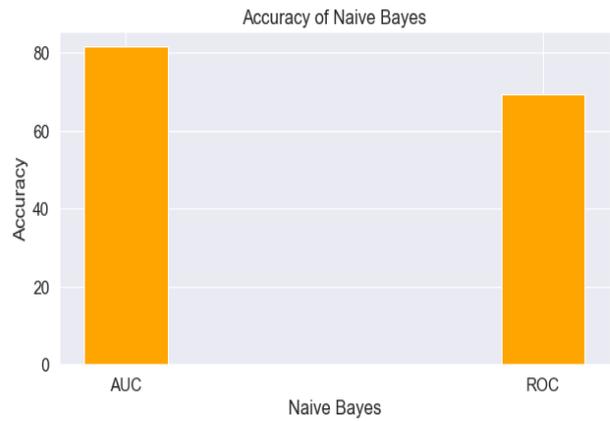


Figure 17. Performance of Naïve Bayes

#### 4.1.4 Support Vector Machine

SVM tested on bread dataset to check the outcome. It has obtained 76.92% accuracy to confidently predict the outcome of bread.

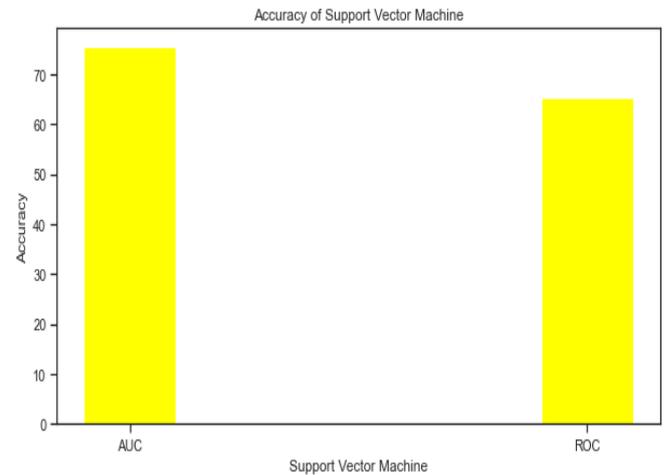
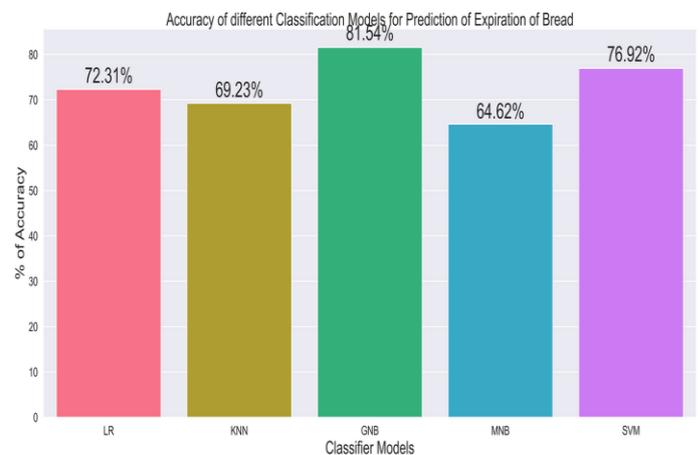


Figure 18. Performance of SVM

Figure below shows the comparative analysis of linear machine learning models. In which Gaussian Naïve Bayes has shown the highest accuracy of 81.54%.



**Figure 23.** Comparative analysis of different linear machine learning models

## 4.2. Comparative Analysis

Due to changes in data, the accuracy of each classifier fluctuates while using balancing techniques on a dataset. The outcomes and comparisons of all approaches utilised in this study are shown in the table below.

Table 4. Comparative Analysis

| Classifier/Algorithm     | Balancing Techniques | Accuracy % |
|--------------------------|----------------------|------------|
| Logistic regression      | SMOTE-TO MEC Links   | 72.31%     |
| SVM                      | SMOTE-TO MEC Links   | 76.92%     |
| KNN                      | SMOTE-TO MEC Links   | 69.23%     |
| Multi Nomial Naive Bayes | SMOTE-TO MEC Links   | 64.62%     |
| Gaussian Naïve Bayes     | SMOTE-TO MEC Links   | 81.54%     |

Above tables shows that Gaussian Naïve Bayes has shown good accuracy after balancing techniques as 81.54%.

## 4. Conclusions

Consumers have expressed an interest in a range of breads, with wheat bread being one of the more popular options available. Quality wheat bread must be made by bakers who are well-liked by the general public in order to be considered a success. Sales will increase significantly if there is a major improvement in the overall quality of the bread. Fungi must feed on the organic matter that they are growing on in order to survive, such as bread or other grains of wheat, in order to survive. Due to the way mushrooms reproduce, they form colonies of spores, which are visible on bread as a result of their reproductive mechanism. Molds such as *Aspergillus* and *Penicillium*, as well as *Mucor* and *Rhizopus*, can be seen growing on bread and other baked items, and they include a variety of species. To make due in a dim, crisp climate with little air dissemination, a contaminated spore should initially distinguish bread and afterward embed its "hyphae" into the bread's surface, a cycle known as hyphal development. In the environment, a mold colony, also known as microbial mycelium, forms as a result of spores from mold rapidly spreading across the environment. First and foremost, in order for organoleptic evaluation to be recognized as a legitimate scientific profession, it is necessary to standardize, rationalize, and link the approach to objective assessment to the method to objective assessment. Organoleptic tests are frequently used in the food industry and other agricultural products to determine the quality of the products being evaluated. With this method of evaluating information, it is feasible to obtain results that are unusually accurate in their accuracy. In some cases, assessment outperforms even the most sensitive of instruments in terms of specific performance. In the food logistics sector, the expiration of a loaf of bread is a very common challenge to deal with. Consumers may become

unwell as a result of consuming fungus bread for a variety of reasons, including the following: Patients may have nausea, diarrhoea, and a range of other medical problems as a result of this condition. Consequently, an intelligent method for determining the current condition of bread is necessary, which will benefit both retailers and consumers. A prototype consisting of an Arduino Nano microcontroller, MQ series sensors for carbon monoxide and carbon dioxide detection, and shopper bags of bread to collect data has been developed as a consequence of this research. In order to determine the current status of the bread in these companies, a number of machine learning algorithms are used to analyze the information collected. One-sided information was acquired by these sensors, resulting in distorted results. The information gathered from sensors is then balanced with the assistance of SMOTE and TO MEC Links, respectively (data balancing techniques). To improve the efficiency of IoT-based datasets, data preparation and feature engineering have been used to them as well. Linear models have been employed to forecast the current condition of the loaf of bread, respectively. Results show that among linear models, the Gaussian Nave Bayes has the highest accuracy with 81.54 percent accuracy, which is the highest accuracy among linear models.

When a dataset is imbalanced, predictive models struggle. Therefore, data balancing techniques (Tomek and SMOTE) have been used to verify that all data points in the dataset are equal in size and weight. Outliers have been removed from the data so that it can be used and interpreted in a more flexible manner. This study also revealed the comparison of several machine learning algorithm-based categorization models to forecast the bread condition at the earliest possible stage. Findings that aren't included here: Classifiers can now be compared in terms of accuracy following the data balancing process. This approach, along with Logistic Regression and Support Vector Machine, as well as K-nearest neighbors, was shown to be 81.54% accurate after a series of comparisons. Bread is a high-carbohydrate food that is widely consumed by members of the general population. To establish the quality of the products being tested, organoleptic tests were often employed in the food sector and other agricultural products. It is possible to acquire exceptionally accurate results by employing this approach of evaluating information. Occasionally, assessment outperforms even the most sensitive of tools in certain respects. Furthermore, this work can be expanded to determine how likely people are to develop diseases associated with fungal infections on food products in the next few years based on factors such as their lifestyle and level of physical activity.

## References

- [1] C. M. Bhatt and J. Nagaraju, "Non-destructive method to estimate the moisture content in Bread using multi-channel electrical impedance spectroscopy," SAS 2009 - IEEE Sensors Appl. Symp. Proc., pp. 56–60, 2009, doi: 10.1109/SAS.2009.4801777.
- [2] M. Morimoto and A. Higasa, "A bread recognition system using RGB-D sensor," 2015 4th Int. Conf. Informatics, Electron. Vision, ICIEV 2015, pp. 1–4, 2015, doi: 10.1109/ICIEV.2015.7334051.

- [3] H. R. Estakhroueiye and E. Rashedi, "Detecting moldy Bread using an E-nose and the KNN classifier," 2015 5th Int. Conf. Comput. Knowl. Eng. ICCKE 2015, pp. 251–255, 2015, doi: 10.1109/ICCKE.2015.7365836.
- [4] R. Attu and M. Terras, "What people study when they study Tumblr," *J. Doc.*, vol. 73, no. 3, pp. 528–554, 2017, doi: 10.1108/jd-08-2016-0101.
- [5] W. J. Arendt and A. I. Arendt, "Aspects of the Breeding Biology of the Cattle Egret (*Bubulcus ibis*) in Montserrat, West Indies, and Its Impact on Nest Vegetation," *Colon. Waterbirds*, vol. 11, no. 1, p. 72, 1988, doi: 10.2307/1521172.
- [6] I. Woodhead, J. Christie, K. Irie, and R. Fenton, "A water content sensor for baked products," 2014 IEEE Sensors Appl. Symp. SAS 2014 - Proc., pp. 20–23, 2014, doi: 10.1109/SAS.2014.6798909.
- [7] C. M. Lavelle, "Gamma ray spectroscopy with Arduino UNO," *Am. J. Phys.*, vol. 86, no. 5, pp. 384–394, 2018, doi: 10.1119/1.5026595.
- [8] B. Popa, I. M. Popescu, D. Popescu, and E. Bobașu, "Real-time monitoring system of a closed oven," *Proc. 2018 19th Int. Carpathian Control Conf. ICC 2018*, pp. 27–32, 2018, doi: 10.1109/CarpathianCC.2018.8399597.
- [9] X. Bao et al., "Yeast Cell Growth Monitoring Using Microwave Measurements Correlated to Optical Absorbance," *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 2018-June, pp. 903–906, 2018, doi: 10.1109/MWSYM.2018.8439349.
- [10] S. N. Chandrasekar, G. Abhishek Rao, V. Sai Muthukumar, S. Venketesh, and R. R. Sarma, "Detection of hotspots in fluorescence imaging of yeast cell model used in neuro-degenerative research," *Proc. 2021 1st Int. Conf. Adv. Electr. Comput. Commun. Sustain. Technol. ICAECT 2021*, pp. 1–4, 2021, doi: 10.1109/ICAECT49130.2021.9392397.
- [11] X. Liu, Y. Cui, T. Zhao, D. Kawashima, H. Obara, and M. Takei, "Quantitative Detection of Living Yeast Fraction from Mixed Living and Dead Cell Solution by Micro Electrical Impedance Spectroscopy," *IEEE Access*, vol. 7, no. 1, pp. 33970–33977, 2019, doi: 10.1109/ACCESS.2019.2902576.
- [12] M. G. Abdallah et al., "Attachment and Detection of Biofouling Yeast Cells Using Biofunctionalized Resonant Sensor Modality," *IEEE Sens. J.*, vol. 21, no. 5, pp. 5995–6002, 2021, doi: 10.1109/JSEN.2020.3040710.
- [13] P. Manikandan and D. Ramyachitra, "An Improved Prorank Algorithm for Detection of Overlapping Protein Complexes using Protein Interaction and Gene Expression Datasets in Yeast," *Proc. - 17th IEEE/ACIS Int. Conf. Comput. Inf. Sci. ICIS 2018*, pp. 110–115, 2018, doi: 10.1109/ICIS.2018.8466500.
- [14] W. Zhang, J. Xu, Y. Li, and X. Zou, "Detecting Essential Proteins Based on Network Topology, Gene Expression Data, and Gene Ontology Information," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 15, no. 1, pp. 109–116, 2018, doi: 10.1109/TCBB.2016.2615931.
- [15] F. Luo, L. Khan, F. Bastani, I. L. Yen, and J. Zhou, "A dynamically growing self-organizing tree (DGSOT) for hierarchical clustering gene expression profiles," *Bioinformatics*, vol. 20, no. 16, pp. 2605–2617, Nov. 2004, doi: 10.1093/bioinformatics/bth292.
- [16] B. Xu et al., "Protein Complexes Detection Based on Global Network Representation Learning," *Proc. - 2018 IEEE Int. Conf. Bioinforma. Biomed. BIBM 2018*, pp. 210–213, 2019, doi: 10.1109/BIBM.2018.8621541.
- [17] G. Li, M. Du, J. Yang, X. Luan, L. Liu, and F. Liu, "Label-Free Rapid Detection of Invasive *S. cerevisiae* Infections with Optically Induced Dielectrophoresis-Based Micromanipulation and Graphene Transistor," *IEEE Sens. J.*, vol. 21, no. 13, pp. 14627–14635, 2021, doi: 10.1109/JSEN.2021.3072387.
- [18] T. Prangemeier, C. Reich, and H. Koepl, "Attention-Based Transformers for Instance Segmentation of Cells in Microstructures," *Proc. - 2020 IEEE Int. Conf. Bioinforma. Biomed. BIBM 2020*, pp. 700–707, 2020, doi: 10.1109/BIBM49941.2020.9313305.
- [19] X. Bao et al., "A planar one-port microwave microfluidic sensor for microliter liquids characterization," *IEEE J. Electromagn. RF Microwaves Med. Biol.*, vol. 2, no. 1, pp. 10–17, 2018, doi: 10.1109/JERM.2018.2807984.
- [20] K. Saboda, R. Moeller, and C. E. Carr, "Enabling Measurement of Darwinian Evolution in Space," *IEEE Aerosp. Conf. Proc.*, vol. 2019-March, pp. 1–10, 2019, doi: 10.1109/AERO.2019.8742164.
- [21] P. Ghosh and A. Kalyanaraman, "FastEtch: A fast sketch-based assembler for genomes," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 16, no. 4, pp. 1091–1106, 2019, doi: 10.1109/TCBB.2017.2737999.
- [22] K. Mallick, S. Mallik, S. Bandyopadhyay, and S. Chakraborty, "A Novel Graph Topology based GO-Similarity Measure for Signature Detection from Multi-Omics Data and its Application to Other Problems," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 5963, no. c, pp. 1–1, 2020, doi: 10.1109/tcbb.2020.3020537.