DEVELOPMENT OF A MACHINE LEARNING BASED SHELF-LIFE PREDICTION MODEL FOR KAJU KATLI

by

ALEENA BIJU (2022-02-047)

SOURAV P (2022-02-050)

NAVYA ANKITA ANAND (2022-02-051)

KERALA AGRICULTURAL UNIVERSITY

DEPARTMENT OF PROCESSING AND FOOD ENGINEERING

KELAPPAJI COLLEGE OF AGRICULTURAL ENGINEERING AND FOOD TECHNOLOGY, TAVANUR-679573,

MALAPPURAM, KERALA, INDIA

2026

DEVELOPMENT OF A MACHINE LEARNING BASED SHELF-LIFE PREDICTION MODEL FOR KAJU KATLI

by

ALEENA BIJU (2022-02-047)

SOURAV P (2022-02-050)

NAVYA ANKITA ANAND (2022-02-051)

PROJECT REPORT

Submitted in partial fulfillment of the requirement for the degree of

BACHELOR OF TECHNOLOGY

in

AGRICULTURAL ENGINEERING

Faculty of Agricultural Engineering and Technology

Kerala Agricultural University

DEPARTMENT OF PROCESSING AND FOOD ENGINEERING

KELAPPAJI COLLEGE OF AGRICULTURAL ENGINEERING AND FOOD TECHNOLOGY, TAVANUR-679573,

MALAPPURAM, KERALA, INDIA

2026

DECLARATION

We hereby declare that this project report entitled **"DEVELOPMENT OF A MACHINE LEARNING BASED SHELF-LIFE PREDICTION MODEL FOR KAJU KATLI"** is a Bonafide record of research work done by us during the course of research and that the report has not previously formed the basis for the award to me of any degree, diploma, associateship, fellowship or other similar title, of any other University or Society.

Place: Tavanur

Date: 30/06/2026

ALEENA BIJU (2022-02-047)

SOURAV P (2022-02-050)

NAVYA ANKITA ANAND (2022-02-051)

CERTIFICATE

Certified that this project entitled **"DEVELOPMENT OF A MACHINE LEARNING BASED SHELF-LIFE PREDICTION MODEL FOR KAJU KATLI"** is a Bonafide record of project work jointly done by Aleena Biju (2022-02-047), Sourav P (2022-02-050), Navya Ankita Anand (2022-02-051) under my guidance and supervision and that it has not previously formed the basis for the award of any degree, diploma, fellowship or associateship to them.

Er. Abhishek S
Guide & Assistant Professor
Dept. of Processing and food Engineering
KCAEFT, Tavanur

Place: Tavanur

Date: 30/06/2026

ACKNOWLEDGEMENT

Words often fall short when attempting to convey our heartfelt gratitude and appreciation. With that sentiment, we, the project team, extend our sincere thanks to our project guide, **Er. Abhishek S**, Assistant Professor for his invaluable guidance, encouragement, and advice throughout this project.

We express our profound gratitude to **Dr. Prince M V,** Head of the Department, Department of Processing and Food Engineering, KCAEFT, Tavanur for his unwavering support and leadership.

We are deeply indebted to **Dr Sruthy P.B.**, Assistant Professor, for her valuable insights and continuous support during the course of this project. We extend our sincere gratitude to **Dr. Jayan P R**, Dean of KCAEFT, for his support and encouragement throughout the project.

Our sincere appreciation goes to **Smt. Geetha T R** (Lab Assistant), KCAEFT, Tavanur for her assistance and continuous support throughout this endeavor.

We are profoundly grateful to our parents for their unconditional love and support, which gave us strength throughout this journey. Our gratitude also extends to the faculty of the library and instructional farm at KCAEFT for their help and cooperation.

Last but certainly not least, we are deeply thankful to our family and friends for their unwavering support, encouragement, and camaraderie, which have been integral to the successful completion of this project. Above all, we bow our heads before **The Almighty** whose grace and blessings have guided and empowered us in every step of this endeavor.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF PLATES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

%	:	Percentage
&	:	And
/	:	Per
+	:	Plus
=	:	Equal to
±	:	Addition or subtraction
°C	:	Degree Celsius
YMC	:	Yeast and Mold Count
SPC	:	Standard Plate Count
ANOVA	:	Analysis of Variance
a_w	:	Water activity
FFA	:	Free Fatty Acid
µm	:	micrometer
ANN	:	Artificial Neural Network
et al.,	:	And others
Etc.,	:	Etcetera
Fig.	:	Figure
g	:	Gram
min	:	Minutes
i.e.	:	That is
KCAEFT	:	Kelappaji College of Agricultural Engineering and Food Technology

rpm	:	Rotation per minute
viz.,	:	Namely
mm	:	millimetre
ml	:	millilitre
m^3	:	meter cube
FDA	:	Food and Drug Administration
GRAS	:	Generally regarded as safe

CHAPTER I
INTRODUCTION

Kaju katli is one of the most popular traditional Indian confectionery products prepared mainly from cashew nuts, sugar, and ghee. The product is widely consumed across India during festivals, celebrations, and special occasions because of its rich taste, attractive appearance, and high nutritional value. However, like many traditional sweet products, kaju katli is highly susceptible to quality deterioration during storage due to microbial growth, moisture migration, lipid oxidation, and environmental conditions. Since the product contains fat-rich cashew components and moisture-sensitive ingredients, its shelf life becomes an important factor influencing consumer safety, product quality, and marketability (Tupe and Ananthanarayan, 2020).

Shelf life can be defined as the duration during which a food product remains safe and acceptable for consumption while retaining its desired sensory, chemical, and microbiological characteristics. Accurate shelf-life prediction is essential for food manufacturers, retailers, and consumers because it helps determine storage conditions, packaging requirements, transportation limits, and expiry dates. Conventional shelf-life estimation methods mainly rely on repeated laboratory testing and observational studies, which are time-consuming, labour-intensive, and expensive. These traditional methods may also fail to capture complex interactions among different quality parameters affecting food spoilage (Wanderlich *et al*., 2023).

Recent developments in machine learning have opened new possibilities in food quality assessment and shelf-life prediction. Machine learning techniques are capable of identifying hidden patterns and relationships from experimental datasets and can generate predictive models with high accuracy. In food science applications, machine learning models are increasingly being used for quality monitoring, spoilage detection, freshness evaluation, and intelligent food processing systems. Among different machine learning algorithms, the Random Forest algorithm is widely preferred because of its robustness, high prediction accuracy, ability to handle nonlinear relationships, and reduced risk of overfitting. The algorithm works by constructing multiple decision trees and combining their outputs to improve prediction reliability.

In the present study, physicochemical and microbiological parameters such as moisture content, water activity, free fatty acid (FFA), yeast and mould count (YMC), and standard plate count (SPC) are used as major indicators influencing the shelf life of kaju katli. Moisture content and water activity play a significant role in determining microbial growth and texture stability. Higher water activity generally accelerates microbial proliferation and spoilage reactions. Similarly, FFA serves as an indicator of lipid degradation and rancidity development in fat-containing products.

Microbiological parameters such as SPC and YMC provide important information regarding bacterial and fungal contamination respectively, which directly affects food safety and storage stability (Lokhande *et al*., 2020).

The experimental data collected through laboratory analysis form the foundation for developing the machine learning model. Since machine learning algorithms perform better with larger and more informative datasets, feature engineering techniques are planned to improve dataset quality and increase the number of effective data points. Feature engineering helps in extracting meaningful relationships from the available data and enhances the predictive capability of the model. The processed dataset is then used to train and test the Random Forest model for shelf-life prediction (Cao. *et al*., 2026).

The significance of this study lies in its practical application in the food industry, especially for small-scale sweet manufacturers and retailers who often face challenges related to product spoilage and uncertain storage stability. An accurate shelf-life prediction system can help reduce food wastage, improve inventory management, enhance consumer confidence, and support quality assurance practices. The model can also reduce dependency on repeated laboratory testing, thereby saving time and operational costs.

Furthermore, this work contributes to the growing integration of artificial intelligence and food technology. The study demonstrates how machine learning can be applied to traditional food products for improving quality monitoring and scientific decision-making. The developed model may later be extended to other Indian confectionery products and bakery items with similar spoilage characteristics. Thus, the research has both academic and industrial relevance in the field of food engineering, food safety, and intelligent food processing systems.

Overall, the proposed shelf-life prediction model for kaju katli represents an innovative and data-driven approach for improving food quality assessment. By combining laboratory-generated quality parameters with machine learning techniques, the study aims to develop a reliable predictive system capable of supporting modern food preservation and management practices.

Objectives

- To analyze changes in moisture content, water activity, free fatty acid (FFA), yeast and mould count (YMC), and standard plate count (SPC) of kaju katli during storage.

- To collect and organize experimental shelf-life data from laboratory analyses and perform feature engineering to improve dataset effectiveness and variability for machine learning applications.

- To develop a Random Forest machine learning model for estimating the shelf life of kaju katli and evaluate the relationship between quality parameters and spoilage behaviour.
- To develop a reliable and cost-effective approach for predicting the storage stability of traditional nut-based confectionery products.

CHAPTER II

REVIEW OF LITERATURE

2.1 KAJU KATLI

Kaju Katli is a traditional Indian nut-based confectionery prepared mainly from cashew nut paste and sugar syrup, valued for its rich flavor, smooth texture, and high nutritional content. Several studies have reported that the quality and shelf life of kaju katli are strongly influenced by factors such as moisture content, water activity, fat oxidation, and microbial growth during storage. Cashew nuts contain a high amount of unsaturated fats, making the product susceptible to lipid oxidation and increase in free fatty acids (FFA), which may lead to rancidity and flavor deterioration over time (Gama *et al*., 2018). Researchers have commonly evaluated shelf life using physicochemical parameters such as moisture content, water activity, peroxide value, and free fatty acid determination, along with microbiological analyses including Standard Plate Count (SPC), yeast, and mold count (YMC). According to the Food Safety and Standards Authority of India (FSSAI), Kaju Katli is officially classified under Food Category 18: Indian Sweets and Snacks & Savories, Category 18.1.3: Dry Fruit and Nut-based Sweets.

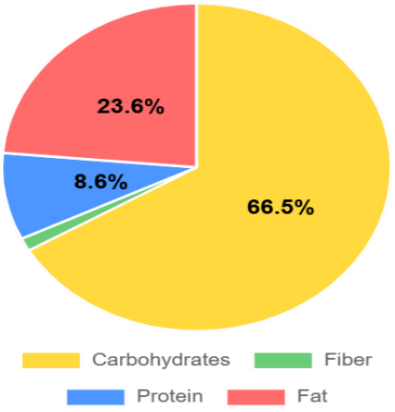
2.1.1 Characteristics and Composition of Kaju Katli

Conventionally, kaju katli is prepared by grinding cashew kernels into a fine powder and mixing it with concentrated sugar syrup under controlled heating conditions until a homogeneous dough-like mass is formed. The cooked mass is subsequently kneaded, rolled into thin sheets, and cut into characteristic diamond-shaped pieces. Variations in preparation methods such as differences in sugar concentration, cooking temperature, kneading time, and incorporation of additional ingredients significantly influence the texture, sensory properties, and storage stability of the final product (Brahmbhatt, 2013)

Nutritionally, kaju katli is considered an energy-dense confection owing to its high carbohydrate content derived from sugar and substantial fat and protein content contributed by cashew nuts. Cashew kernels contain approximately 48.3% fat, 21.3 g/100 g protein, and 20.5 g/100 g carbohydrates, with a high proportion of unsaturated fatty acids such as oleic and linoleic acids (Rico *et al*., 2016).

They are rich sources of unsaturated fatty acids, proteins, and essential minerals such as magnesium, zinc, copper, and iron along with antioxidant compounds and vitamins such as vitamin E. Recent studies have also reported the development of modified formulations including low-sugar and sugar-free variants to improve nutritional quality and consumer acceptability. The presence of

these nutrients enhances the nutritional value of the product; however, the relatively high fat content also increases susceptibility to lipid oxidation and rancidity during storage. In addition, the moisture content and water activity of kaju katli influence microbial growth, texture deterioration, and overall shelf stability. Therefore, monitoring physicochemical and microbiological characteristics becomes important for shelf-life prediction. (Patel *et al*., 2021).



NUTRIENT	AMOUNT
Calories	120 kcal
Protein	2 g
Carbohydrates	15.5 g
Fiber	0.3 g
Sugars	10 g
Fat	5.5 g
Saturated Fat	1 g
Copper	0.5 mg
Magnesium	18 mg
Phosphorus	30 mg

Fig 2.1 Nutritional information of Kaju katli

2.1.2. Manufacturing Process of Kaju Katli

Conventional preparation involves grinding cashew kernels into a fine powder followed by mixing with concentrated sugar syrup under controlled heating conditions. The mixture is continuously stirred until a soft dough-like consistency is obtained, after which it is kneaded, rolled into thin sheets, and cut into diamond-shaped pieces. Additional ingredients such as ghee, cardamom, milk solids, and edible silver foil may also be incorporated depending on regional and commercial practices.

Several researchers have studied the formulation of kaju katli with respect to ingredient proportion and sensory acceptability. Tupe and Ananthanarayan (2021) reported that a formulation containing 65% cashew nut and 35% sugar produced optimum sensory quality and consumer acceptability. The study concluded that this ratio provided desirable sweetness, texture, cohesiveness, and mouthfeel while maintaining the characteristic flavor of cashew nuts.

The preparation involves soaking and grinding cashew nuts into a fine paste or powder, followed by cooking it with sugar syrup of appropriate consistency (typically two-thread or 110–115 °C) until a dough-like mass is formed, which is then rolled and cut into diamond shapes.

The proportion of cashew powder to sugar is an important factor influencing the sensory quality, nutritional characteristics, texture, and storage stability of Kajukatli. Tupe and

Ananthanarayan (2020) evaluated different levels of cashew nut and sugar in Kajukatli and identified a formulation containing 65% cashew nut and 35% sugar as the optimum combination based on sensory analysis. The study also highlighted the importance of controlling sugar and other ingredients in Kajukatli, as the product is susceptible to adulteration or dilution with inexpensive ingredients. Patel *et al*. (2021) developed a sugar-free Kajukatli by replacing sucrose with suitable sweeteners and bulking agents and reported that the choice of bulking agent significantly influenced the texture and taste of the product. Cashew kernels themselves are nutritionally rich, containing substantial amounts of fat and protein, which contribute to the nutritional and functional characteristics of cashew-based products (Rico *et al*., 2016). The high lipid content of cashew also makes oxidation and storage conditions important considerations for product quality, as tree nuts are susceptible to oxidative deterioration during storage (Gama *et al*., 2018). Furthermore, Sharma *et al*. (2017) reported that standardized Kajukatli had an acceptable shelf life of approximately 9 days at room temperature and more than 30 days under refrigerated conditions, demonstrating the influence of storage temperature on the quality and shelf life of the product. Justification for selecting the 65:35 ratio: The 65:35 cashew powder-to-sugar ratio was preferred in the present study because:

i. Compiles with FSSAI premium-grade nut sweet specifications.
ii. Provides the highest hedonic acceptability as reported in published literature.
iii. Ensures a balanced macronutrient profile with appreciable protein and healthy fat content.
iv. Yields desirable rheological and textural properties (soft, pliable, non-greasy).
v. Maintains the authentic cashew flavour without excessive sweetness, aligning with current consumer preference for reduced-sugar traditional sweets (Kumbhar *et al*., 2018; Patil *et al*., 2020).

2.2. SHELF-LIFE OF CONFECTIONERY PRODUCTS

2.2.1. Definition of Shelf-life

According to the FAO/WHO Codex Alimentarius, shelf life is defined as the estimated period during which a food product maintains its microbiological safety and sensory qualities under specific, stated storage conditions. The shelf life of confectionery products depends on factors such as moisture content, water activity, fat composition, packaging, storage temperature, and microbial stability (Bauer *et al*., 2022).

The shelf life of kaju katli is primarily affected by physicochemical deterioration and microbial spoilage. Due to its fat-rich composition, the product is susceptible to oxidative rancidity during storage. Lipid oxidation results in the formation of free fatty acids and peroxide compounds,

leading to off-flavors, unpleasant odor, and quality deterioration. Moisture migration and environmental humidity further influence texture softening or hardening during storage.

Microbial spoilage is another important factor limiting shelf life. Increased water activity and improper storage conditions favor the growth of bacteria, yeast, and molds. Yeasts and molds are especially important spoilage organisms in confectionery products stored at ambient conditions. Elevated microbial load can reduce product safety and consumer acceptability (Tarlak, 2023).

2.2.2. Factors Affecting Shelf-life of Confectionery Products

The shelf life of confectionery products is influenced by a combination of physicochemical, microbiological, and environmental factors. Traditional Indian confectioneries such as kaju katli are particularly susceptible to quality deterioration because of their high fat content, intermediate moisture level, and nutrient-rich composition. During storage, various reactions occur that affect product safety, sensory quality, and consumer acceptability. Previous studies have demonstrated that storage temperature, moisture migration, water activity, lipid oxidation, and microbial growth are among the most important determinants of shelf stability in confectionery products (Fellows, 2022; Fennema, 2017).

The storage temperature significantly influenced textural characteristics such as cohesiveness, adhesiveness, and springiness. Fluctuating environmental conditions accelerated quality deterioration and reduced product stability. Similar observations have been reported for cashew-based sweets where elevated storage temperatures accelerated oxidation, microbial growth, and texture degradation.

Being a high-fat, intermediate-moisture confectionery, kaju katli is susceptible to multiple modes of deterioration during storage. The major spoilage mechanisms reported in the literature include:

Moisture Content and Moisture Migration

Moisture content is one of the most critical factors governing the shelf life of confectionery products. Moisture migration occurs when water moves between different regions of a product or between the product and the surrounding environment due to differences in water vapour pressure. Such migration can alter texture, flavour, and overall product quality.

In products such as kaju katli, excessive moisture absorption may result in stickiness, softening, and increased susceptibility to microbial spoilage. Conversely, excessive moisture loss can lead to hardening, brittleness, and reduced consumer acceptability. Moisture changes also influence the rate of lipid oxidation and enzymatic reactions occurring within the product matrix.

According to Fellows (2022), moisture migration is one of the primary causes of quality

deterioration in intermediate-moisture foods. The stability of confectionery products therefore depends on maintaining appropriate moisture levels throughout storage. Packaging materials with adequate moisture barrier properties are often employed to minimize undesirable moisture transfer and extend shelf life. The slight deviation in the moisture content can result in softer or harder texture.

Water Activity (a_w)

Water activity (a_w) is considered a more reliable indicator of microbial stability than total moisture content because it measures the availability of free water for microbial growth and chemical reactions. Water activity values range from 0 to 1, where pure water has an aw of 1.0.

Most pathogenic and spoilage bacteria require aw values greater than 0.90 for growth, whereas yeasts and moulds can grow at significantly lower values. Xerophilic moulds and osmophilic yeasts are capable of surviving at aw levels commonly encountered in confectionery products. Beuchat (2002) reported that mould growth can occur at aw values as low as 0.60–0.65, depending on environmental conditions and product composition.

Kaju katli generally exhibits water activity values ranging between 0.55 and 0.75. Although these levels inhibit the growth of many bacteria, they may still permit the proliferation of certain yeasts and moulds during prolonged storage. Changes in water activity during storage can therefore directly influence microbial stability, texture, and shelf life.

Lipid Oxidation and Oxidative Rancidity

Lipid oxidation is one of the most important deterioration mechanisms affecting cashew-based confectionery products. Cashew kernels contain approximately 44–48% lipids, predominantly unsaturated fatty acids such as oleic acid and linoleic acid (Lima *et al*., 2012). These unsaturated fatty acids are highly susceptible to oxidation when exposed to oxygen, light, and elevated temperatures. The oxidation process begins with the formation of lipid free radicals, followed by the production of hydroperoxides. These unstable compounds subsequently decompose into aldehydes, ketones, alcohols, and other volatile compounds responsible for rancid odours and off-flavours (Frankel, 2014).

Trox *et al*., (2010) reported significant increases in peroxide value and free fatty acid content during storage of cashew-based products under ambient conditions. Similarly, Yadav *et al*., (2019) observed a progressive increase in FFA content in cashew-containing confectionery products stored at room temperature. Elevated FFA levels are commonly associated with loss of flavour quality and reduced consumer acceptance. Since kaju katli contains a substantial proportion of cashew lipids, monitoring FFA content provides an effective indicator of oxidative deterioration and shelf-life

progression.

Hydrolytic Rancidity

Apart from oxidative deterioration, hydrolytic rancidity also contributes significantly to quality loss during storage. Hydrolysis of triglycerides occurs in the presence of moisture and lipolytic enzymes, resulting in the formation of free fatty acids and glycerol.

Residual moisture within the product and enzymes originating from raw materials or microorganisms can accelerate hydrolytic reactions. Frankel (2014) reported that hydrolytic rancidity is particularly important in foods containing moderate moisture levels because water facilitates the cleavage of ester bonds within triglycerides.

The increase in free fatty acid concentration associated with hydrolytic rancidity contributes to undesirable flavour changes and serves as a useful indicator of product deterioration.

Microbial Spoilage

Microbial spoilage is one of the principal factors limiting the shelf life of confectionery products. Although kaju katli contains a high concentration of sugar, which lowers water activity and inhibits the growth of many microorganisms, it remains susceptible to spoilage by osmophilic yeasts, xerophilic moulds, and certain bacteria capable of surviving under reduced water activity conditions. The risk of microbial growth increases during storage due to moisture migration, improper packaging, temperature fluctuations, and post-processing contamination.

Microbial contamination in confectionery products can originate from raw materials such as cashew kernels, sugar, processing equipment, packaging materials, and handling practices. Cashew nuts are particularly vulnerable to fungal contamination during harvesting, drying, storage, and transportation. Once introduced into the product matrix, these microorganisms may proliferate during storage if environmental conditions become favourable.

- **Standard Plate Count (SPC)**

  Standard Plate Count (SPC), also referred to as Total Plate Count (TPC) or Aerobic Plate Count (APC), is widely used as an indicator of the overall microbiological quality of food products. SPC estimates the population of viable aerobic mesophilic microorganisms present in a sample and serves as an indicator of hygienic processing and storage conditions.

  An increase in SPC during storage generally indicates microbial proliferation and progressive deterioration of product quality. Elevated SPC values may result in undesirable changes in flavour, odour, appearance, and texture. Although SPC does not specifically identify pathogenic microorganisms, it provides a useful measure of overall microbial load and product stability.

  Several studies on traditional Indian confectionery products have reported gradual increases

in SPC during storage under ambient conditions. The increase is often attributed to contamination from raw materials and environmental exposure during processing and storage. Refrigerated storage has been shown to significantly reduce microbial growth rates and delay spoilage.

- **Yeast and Mold Count (YMC)**

Yeast and Mold Count (YMC) is a critical microbiological parameter for evaluating the shelf life of confectionery products because fungi are capable of growing at lower water activity values than most bacteria. High-sugar foods such as kaju katli may therefore remain susceptible to fungal spoilage even when bacterial growth is limited.

Yeasts commonly associated with confectionery spoilage include:

- *Zygosaccharomyces rouxii*
- *Saccharomyces cerevisiae*
- *Candida tropicalis*
- *Debaryomyces hansenii*

These yeasts are osmophilic and can tolerate elevated sugar concentrations. Their growth may result in fermentation, gas production, undesirable flavours, and changes in texture.

Moulds are frequently the predominant spoilage organisms in nut-based confectionery products. Common mould genera reported in confectionery and nut products include:

- *Aspergillus flavus*
- *Aspergillus niger*
- *Aspergillus parasiticus*
- *Penicillium chrysogenum*
- *Penicillium expansum*
- *Eurotium amstelodami*
- *Rhizopus stolonifer*
- *Mucor* spp.

These moulds may produce visible fungal growth, discoloration, off-flavours, texture defects, and in some cases mycotoxins that pose serious health risks.

- **Aspergillus Species and Aflatoxin Concerns**

Among the fungal contaminants associated with nut-based confectionery products, *Aspergillus flavus* and *Aspergillus parasiticus* are of particular concern due to their ability to produce aflatoxins. Aflatoxins are toxic secondary metabolites that are highly stable and have been classified as human carcinogens.

Cashew kernels and other nuts may become contaminated with *Aspergillus* species during storage under warm and humid conditions. The isolation of *Aspergillus flavus* from spoiled cashew-based

sweets, highlighting the importance of controlling fungal contamination throughout production and storage. The presence of these fungi may significantly reduce product shelf life and compromise food safety.

Hosen *et al*., (2025) reported that confectionery products with intermediate moisture content can support the growth of spoilage fungi during storage. Common mould genera associated with nut-based confectioneries include *Aspergillus*, *Penicillium*, and *Eurotium*. These microorganisms can cause visible spoilage, off-flavours, discoloration, and potential mycotoxin production.

- **Microbial Spoilage in Other Nut-Based Confectionery Products**

Similar microbial spoilage patterns have been reported in various nut-based confectionery products such as peanut chikki, almond burfi, pista sweets, nut bars, pralines, marzipan, and chocolate-coated nut products. Studies have identified *Aspergillus*, *Penicillium*, *Eurotium*, and *Rhizopus* species as dominant spoilage organisms in these products. The susceptibility of nut-based confectioneries to fungal spoilage is primarily attributed to their lipid-rich composition and intermediate moisture levels.

Storage at ambient temperatures generally results in faster microbial growth compared with refrigerated conditions. Several studies have reported significantly lower SPC and YMC values in refrigerated products due to reduced metabolic activity and slower microbial proliferation.

- **FSSAI Microbiological Considerations**

The Food Safety and Standards Authority of India (FSSAI) specify microbiological criteria for various food categories and recommends routine monitoring of aerobic plate count and yeast and mould count as indicators of hygienic quality and product stability. FSSAI microbiological testing manuals include standard procedures for determination of Total Plate Count and Yeast and Mold Count in confectionery products.

For high-sugar and low-moisture foods, excessive increases in SPC and YMC during storage are generally considered indicators of deterioration and reduced shelf stability. Yeast and mould growth is often regarded as the primary microbiological factor limiting the shelf life of nut-based confectionery products because fungal species can survive and proliferate at lower water activity levels than bacteria. Consequently, monitoring SPC and YMC provides valuable information regarding microbial safety, product quality, and remaining shelf life.

Since kaju katli contains substantial amounts of cashew solids, lipids, and carbohydrates, microbial spoilage represents a critical quality concern during storage. Increases in SPC indicate progressive microbial proliferation, while increases in YMC reflect the growth of spoilage yeasts and moulds capable of surviving under reduced water activity conditions.

Effect of Storage Temperature

Storage temperature is the primary environmental factor controlling the shelf life of nut-based confectionery by dictating the kinetics of chemical, physical, and microbiological degradation. According to the Arrhenius law, the rate of chemical reactions increases exponentially with temperature. Elevated temperatures supply the activation energy required to accelerate destructive pathways chiefly lipid oxidation, non-enzymatic browning (Maillard reaction), and moisture migration. In contrast, low-temperature storage (refrigeration at 4-7°C) slows down molecular kinetic energy, effectively preserving the structural and sensory profile of the food matrix.

Impact on Lipid Oxidation and Rancidity

Cashew nuts (*Anacardium occidentale*), the structural base of *kaju katli*, contain approximately 40-45% lipids, predominantly rich in unsaturated fatty acids like oleic acid and linoleic acid.

High storage temperatures lower the energy barrier for lipid auto-oxidation. This accelerates the degradation of primary hydroperoxides into volatile secondary oxidation products such as hexanal and malondialdehyde (MDA) (Raisi *et al*., 2015). Research exploring the shelf life of temperate nuts demonstrated that matrices stored at or above 20°C exhibited significantly elevated Peroxide Values (PV) and Thiobarbituric Acid Reactive Substances (TBARS) compared to those under cold storage at 4°C (Sharma *et al*., 2017). This acceleration is tied to elevated endogenous lipoxygenase and lipase enzyme activity triggered by ambient heat. Similarly, an Accelerated Shelf-Life Testing (ASLT) study on cashew-based confections verified that lipid rancidity adheres tightly to first-order reaction kinetics, heavily governed by storage temperature (Raisi *et al*., 2015).

Moisture Migration and Textural Kinetics

Kaju katli requires a delicate balance of moisture (typically 10-15%) to maintain its characteristic soft, pliable texture (Sharma *et al*., 2017). Temperature directly modulates the thermodynamic water activity (a_w) of the product. Elevated temperatures increase the vapour pressure of water within the confectionery matrix, provoking moisture migration toward the environment or packaging headspace. This often causes sugar crystallization and protein-lipid cross-linking, turning the product hard and crumbly. A dedicated storage study on standardized *kaju katli* demonstrated that samples stored at room temperature (30 ± 2°C, 65% RH) underwent intense textural hardening within 9 days due to rapid moisture loss and oil separation (Sharma *et al*., 2017). Conversely, keeping the sweet under refrigeration (7 ± 2°C, 90% RH) effectively minimized vapour pressure differentials, maintaining textural tenderness and acceptable sensory properties for over 30 days (Sharma *et al*., 2017).

Inhibition of Microbiological Deterioration

Although the relatively high sugar content of *kaju katli* exerts an osmotic pressure that limits broad bacterial spoilage, its intermediate moisture level leaves it vulnerable to xerophilic molds and osmophilic yeasts (Sharma *et al*., 2017). The specific growth rate (μ) of spoilage microorganisms is highly temperature-dependent, fitting the classic Ratkowsky square-root model. Ambient conditions compress the lag phase of molds and total aerobic bacteria, yielding exponential colony proliferation. Empirical data verified that room-temperature storage (30°C) leads to a rapid surge in Total Plate Counts (TPC) and Yeast & Mold Counts (YMC) beyond permissible food safety limits after just 9 days (Sharma *et al*., 2017). Refrigeration at 7°C effectively halts microbial enzymatic pathways, extending the microbiological stability safely past 4 weeks (Sharma *et al*., 2017; Mishra *et al*., 2021).

2.3. WATER ACTIVITY (a_w) AND SHELF-LIFE OF KAJU KATLI

Water activity (a_w) is one of the most important factors governing the shelf life and microbiological stability of food products. It is defined as the ratio of the vapour pressure of water in a food system to the vapour pressure of pure water at the same temperature. Unlike moisture content, which represents the total amount of water present in a product, water activity measures the amount of free or available water that can participate in chemical reactions and support microbial growth. Consequently, water activity is considered a more reliable predictor of microbial stability than moisture content alone.

According to previous storage studies done on various kaju katli samples, the water activity analysis revealed that the moisture content and water activity reported to be important factors to affect the keeping quality and texture of the product. The slight deviation in the moisture content can result in softer or harder texture. The change in water activity (a_w) may favor growth of bacteria and fungi which requires minimum water activity of 0.91 and 0.7 respectively for their support of growth (Parmar, 2012; Macwan 2012). water activity of the stored samples decreases as storage period progressed. The initial value was 0.78 and it remains higher on 12th day of storage then safe value of 0.7. Samples were analyzed in this study using a Novasina water activity meter.

Changes in water activity significantly influence the quality and shelf life of confectionery products. An increase in water activity provides greater availability of free water for microbial metabolism, thereby promoting the growth of spoilage microorganisms such as bacteria, yeasts, and moulds. Consequently, elevated water activity levels may accelerate microbial spoilage, reduce product safety, and shorten shelf life. Conversely, a decrease in water activity restricts microbial growth and enhances microbiological stability; however, excessive reduction in water activity can adversely affect the physical properties of the product. Lower water activity is often associated with

moisture loss, resulting in increased hardness, reduced cohesiveness, loss of softness, and undesirable textural changes. Such alterations may negatively impact consumer acceptability even when microbial spoilage is minimal. Therefore, maintaining an optimum water activity level is essential to preserve both the microbiological safety and desirable sensory characteristics of kaju katli during storage.

Water activity was expected to be a highly informative predictive feature because changes in water activity are often accompanied by corresponding changes in microbial counts (SPC and YMC), moisture migration, and textural properties, all of which contribute significantly to shelf-life reduction in nut-based confectionery products.

2.4. MOISTURE CHANGES DURING STORAGE

Moisture content is one of the primary quality attributes influencing the shelf life of confectionery products. During storage, moisture migration may occur between the product and surrounding environment, leading to changes in texture and acceptability. A progressive decrease in moisture content during storage of Indian confectionery products, resulting in increased hardness and reduced cohesiveness. Similar reductions in moisture content have been reported for nut-based sweets stored under ambient conditions.

Previous studies on effects of storage conditions on the shelf-life of kaju katli have shown that moisture content of samples stored in room condition decreased from initial level of 9.62 to 6.73% (w/w basis) over a period of 12 days. For refrigerated sample, at the end of thirty days of storage moisture content of the product was 7.59%.

Sachdeva and Rajhoria (1982) also reported decrease from initial 23.84% moisture content of plain burfi during storage at 30˚C to 15.44% at the end of 10 days. The change reported due to loss of product's moisture, thereby decrease in water content and activity.

Moisture loss in Kaju Katli during storage primarily occurs due to moisture migration from the product to the surrounding environment, especially when it is exposed to lower relative humidity or poorly sealed packaging. Since the product is a sugar-cashew matrix, water is not strongly bound, so it gradually evaporates or redistributes within the matrix over time. This reduction in moisture directly affects texture by increasing hardness and reducing softness and chewiness, as lower water content strengthens intermolecular bonding in the sugar matrix and reduces plasticity. As a result, the product transitions from a smooth, pliable texture to a brittle or grainy one, which is considered a major quality defect in sensory evaluation.

Moisture content is closely linked with water activity (a_w), which represents the availability of free water for microbial growth and chemical reactions. Even if total moisture decreases slightly,

a small change in water activity can significantly influence shelf stability because lipid oxidation, Maillard browning, and microbial growth are highly water-activity dependent. In Kaju Katli, decreasing moisture generally lowers water activity, but uneven distribution of water can still create localized regions favourable for spoilage reactions.

From a machine learning perspective, moisture content and water activity are strong predictive features for shelf-life modelling because they act as primary drivers of both physicochemical degradation (texture hardening, lipid oxidation) and microbial stability. Their time-dependent behavior captures the underlying degradation kinetics of the product, making them highly informative variables for predicting remaining shelf life compared to indirect sensory parameters.

2.5. LIPID OXIDATION AND FREE FATTY ACID (FFA)

Lipid oxidation is one of the major causes of quality deterioration in nut-based confectionery products. Cashew kernels contain approximately 44-48% lipids, with oleic acid and linoleic acid constituting the predominant fatty acids. Although these unsaturated fatty acids contribute desirable nutritional and sensory properties, they are highly susceptible to oxidation during storage, particularly under conditions of elevated temperature and oxygen exposure (Lima *et al*., 2012).

Lipid oxidation occurs through a free-radical chain reaction involving initiation, propagation, and termination stages. During this process, unsaturated fatty acids react with oxygen to form hydroperoxides, which are unstable intermediates. Subsequent decomposition of hydroperoxides produces secondary oxidation products such as aldehydes, ketones, alcohols, and hydrocarbons. These compounds are responsible for rancid flavours, undesirable odours, and deterioration of sensory quality (Frankel, 2014).

Apart from oxidative rancidity, hydrolytic rancidity also contributes significantly to quality degradation. In the presence of moisture and lipolytic enzymes, triglycerides undergo hydrolysis, resulting in the formation of free fatty acids and glycerol. This reaction increases acidity and produces undesirable flavour characteristics. Free fatty acid content therefore serves as an important indicator of lipid deterioration in fat-rich food products.

Several studies on cashew kernels and nut-based confectionery products have reported progressive increases in free fatty acid levels during storage. The increase is generally more pronounced under ambient storage conditions due to accelerated oxidation and hydrolysis. Elevated free fatty acid levels have been associated with loss of flavour quality, reduced consumer acceptability, and shortened shelf life.

Because free fatty acid content reflects both oxidative and hydrolytic deterioration processes, it is widely recognized as a reliable indicator of quality loss in nut-based confectionery

products. Therefore, FFA was selected as one of the key shelf-life indicators in the present study and was incorporated as an input variable in the machine learning model developed for predicting the remaining shelf life of kaju katli.

2.6. MICROBIOLOGICAL CRITERIA (SPC AND YMC)

Microbiological quality is a critical determinant of food safety and shelf life. The growth of microorganisms during storage can lead to spoilage, undesirable sensory changes, reduced product acceptability, and potential health hazards. Standard Plate Count (SPC) and Yeast and Mold Count (YMC) are among the most widely used microbiological indicators for assessing product quality and storage stability. SPC represents the total viable aerobic microbial population present in a product, whereas YMC specifically quantifies fungal contamination arising from yeasts and moulds.

Previous storage studies on kaju katli assessed microbial quality using standard microbiological procedures. Standard Plate Count was determined using Nutrient Agar, while Yeast and Mold Count was evaluated using acidified Potato Dextrose Agar. These methods provided valuable information regarding microbial proliferation and product stability during storage. Microbiological analyses were performed periodically to monitor the progression of spoilage and determine the microbiological shelf life of the product.

Several microorganisms have been associated with spoilage of nut-based confectionery products. Common bacterial contaminants include *Bacillus subtilis*, *Bacillus cereus*, *Micrococcus luteus*, and *Staphylococcus aureus*. These organisms may enter the product through raw materials, processing environments, packaging materials, or handling operations. Although the low water activity of confectionery products limits bacterial growth, microbial proliferation may occur when favourable storage conditions exist.

Yeasts frequently associated with spoilage of high-sugar confectionery products include *Zygosaccharomyces rouxii* and *Saccharomyces cerevisiae*. These organisms are highly osmotolerant and capable of surviving in environments containing elevated sugar concentrations. Their growth may result in fermentation, gas production, flavour defects, and texture changes. Mould species commonly isolated from nut-based confectionery products include *Aspergillus flavus*, *Aspergillus niger*, *Penicillium chrysogenum*, and *Eurotium amstelodami*. These fungi can proliferate at relatively low water activity values and often represent the primary spoilage organisms in confectionery products.

Among the moulds associated with confectionery spoilage, *Aspergillus flavus* is of particular concern because of its ability to produce aflatoxins. Aflatoxins are toxic secondary metabolites that have been classified as carcinogenic to humans. Contamination of nuts by *Aspergillus* species may

occur during harvesting, drying, storage, and transportation, thereby posing significant food safety concerns in nut-based products.

For microbiological quality assessment, FSSAI reference criteria for similar traditional sweet products specify an acceptable Standard Plate Count of approximately **4.40 log CFU/g** and a maximum permissible level of **4.88 log CFU/g**. Similarly, the acceptable Yeast and Mold Count is approximately **1.00 log CFU/g**, while the maximum permissible level is **2.0 log CFU/g**. Exceeding these limits indicates deterioration in microbiological quality and reduced shelf stability.

Numerous studies have demonstrated that microbial counts increase progressively during storage, particularly under ambient conditions. Elevated temperatures and increased water activity accelerate microbial proliferation, resulting in higher SPC and YMC values and consequently shorter shelf life. Refrigerated storage generally slows microbial growth and extends product stability by reducing the rate of microbial metabolism.

Because microbial spoilage is one of the principal causes of shelf-life termination in confectionery products, SPC and YMC were selected as key microbiological indicators in the present study. Their ability to reflect product safety and quality deterioration also made them valuable predictor variables for the machine learning model developed for shelf-life prediction of kaju katli.

2.7. MACHINE LEARNING IN FOOD SHELF-LIFE PREDICTION

Machine learning (ML) has emerged as a powerful tool for predicting food shelf life by identifying complex relationships between physicochemical, microbiological, sensory, and environmental variables. Unlike conventional statistical approaches, machine learning algorithms can process large datasets and model nonlinear interactions among quality parameters, resulting in more accurate shelf-life predictions. The application of machine learning in food science has gained considerable attention due to its potential to improve food safety, reduce product losses, and minimize food waste throughout the supply chain (Tarlak, 2023).

Several machine learning techniques have been successfully applied for shelf-life prediction in various food products. Among these, Artificial Neural Networks (ANN), Support Vector Machines (SVM), Decision Trees, Random Forests, and Deep Learning models have demonstrated promising predictive performance. According to Arteaga-Cabrera *et al*., (2025), ANN-based models are capable of accurately predicting food deterioration by learning complex relationships between storage conditions and quality attributes. Their review highlighted that machine learning approaches often outperform traditional kinetic and regression models in predicting food shelf life because they can simultaneously consider multiple interacting variables such as temperature, humidity, microbial growth, and chemical changes.

Recent advancements in artificial intelligence have further expanded the application of predictive modelling in food systems. Li *et al*., (2025) reported that deep learning models have shown high predictive accuracy for shelf-life estimation in a wide range of food products, including fruits, vegetables, dairy products, meat products, and processed foods. The authors emphasized that machine learning models can integrate physicochemical, microbiological, and environmental data to provide more reliable shelf-life predictions compared to conventional methods.

Similarly, studies reviewed by Tarlak (2023) demonstrated that predictive models incorporating microbial growth data can effectively estimate microbial shelf life and spoilage progression. Such approaches allow food manufacturers to predict product deterioration before visible spoilage occurs, thereby improving quality control and inventory management. Predictive microbiology combined with machine learning has become an important research area because microbial growth is one of the primary factors limiting food shelf life.

Artificial intelligence-based shelf-life prediction systems have also been applied using quality indicators such as moisture content, water activity, pH, microbial counts, temperature, and storage duration. Recent reviews indicate that machine learning models are capable of predicting quality deterioration with high accuracy by analysing complex interactions among these variables. Such models provide significant advantages over traditional shelf-life studies, which are often time-consuming, labour-intensive, and costly.

2.7.1. Introduction to Machine Learning

Machine learning (ML) is a branch of artificial intelligence that enables computers to learn patterns from historical data and make predictions without being explicitly programmed. In food science, machine learning has gained significant attention for predicting quality changes, spoilage progression, and shelf life of food products. Traditional shelf-life determination methods require lengthy storage studies and repeated laboratory analyses, whereas machine learning models can utilize existing physicochemical, microbiological, and environmental data to estimate remaining shelf life more efficiently and accurately (Arteaga-Cabrera *et al*., 2025).

2.7.2. Common Workflow in Machine Learning Projects

The development of a machine learning framework for food shelf-life prediction follows a systematic, data-driven workflow designed to map complex degradation dynamics into accurate temporal predictions. The process initiates with Data Acquisition and Preprocessing, where heterogeneous data streams-encompassing environmental parameters (temperature, relative humidity), microbial loads, chemical biomarkers (pH, lipid oxidation), and sensory scores-are collected, cleaned of anomalies, and normalized. This is followed by Feature Engineering and Selection, which isolates the most critical kinetic descriptors of spoilage, often integrating traditional

kinetic equations (such as the Arrhenius model) with statistical feature ranking to reduce dimensionality. During the Model Selection and Training phase, the curated dataset is partitioned into training and validation subsets to optimize various candidate algorithms, ranging from classical regressors (e.g., Random Forests, Support Vector Machines) to advanced Artificial Neural Networks (ANNs). The predictive accuracy and generalization capabilities of these models are then rigorously quantified in the Model Evaluation phase using standard statistical metrics, including Root Mean Squared Error (RMSE) and the Coefficient of Determination (R^2). Ultimately, the optimized pipeline concludes with Deployment, where the finalized model is integrated into cloud platforms or Internet-of-Things (IoT) smart packaging systems to enable real-time, continuous freshness monitoring across the food supply chain.

1) **Data Acquisition:** The first step involves data collection, where relevant quality parameters are measured during storage. These may include physicochemical attributes such as moisture content, water activity, pH, colour, texture, and free fatty acid content, along with microbiological indicators such as standard plate count (SPC) and yeast and mold count (YMC). The collected data are then organized into a structured dataset suitable for analysis.
2) **Data Preprocessing:** The second step is data preprocessing, which involves handling missing values, removing inconsistencies, normalizing variables if necessary, and identifying outliers. Proper preprocessing improves data quality and enhances model performance. After preprocessing, the dataset is divided into training and testing subsets. The training dataset is used to develop the predictive model, while the testing dataset is used to evaluate its performance on unseen data.
3) **Feature Selection and Model Development:** The third step involves feature selection and model development. Feature selection identifies the most informative variables influencing shelf life. Commonly used machine learning algorithms include Linear Regression, Support Vector Machines (SVM), Artificial Neural Networks (ANN), Decision Trees, Random Forests, and Deep Learning models. The selected algorithm learns relationships between input variables and the target variable, which in shelf-life studies is typically the remaining shelf life or storage duration.
4) **Model Evaluation and Validation:** The fourth step is model evaluation and validation. Performance indicators such as coefficient of determination (R^2), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) are commonly used to assess prediction accuracy. Models demonstrating high R^2 values and low prediction errors are considered suitable for practical applications. Finally, the validated model is deployed

for predicting shelf life of new samples using measured quality parameters.

Regression Parameters Used in Analysis of Machine Learning Models

The results of the machine learning models are evaluated using various important regression parameters discussed as follows. These metrics offer insights into the magnitude and distribution of prediction errors.

- Mean square error (MSE): The mean or average of the squared discrepancies between the anticipated and expected target values in the dataset is used to calculate the MSE. It is calculated as

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(x_a - x_p\right)^2$$

where x_a is the actual value of the i_{th} object and x_p is the predicted value of the i_{th} object for n samples.

- Root means square error (RMSE) The square root of the mean of the squared discrepancies between the anticipated and expected values in the dataset is used to calculate the MSE. It is calculated as

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_a - x_p\right)^2}$$

where x_a is the actual value of the i_{th} object and x_p is the predicted value of the i_{th} object for n samples.

- Mean absolute error (MAE) It is calculated as the average of the absolute difference between the target and predicted values of the object. It is defined as

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|x_a - x_p\right|$$

where x_a is the actual value of the i_{th} object and x_p is the predicted value of the i_{th} object for n samples.

- R square (R^2) R-squared indicates how much the variation of one variable contributes to the variance of the second. In other words, it calculates the percentage of the variance that can be accounted for by the independent variables. It is calculated as

$$R^2 = 1 - \frac{SSE}{SST}$$

where SSE is the sum of the square of the difference between the actual value and the predicted value, and SST is the sum of the square of the difference between the actual value and the mean of the actual value. They are calculated as

$$SSE = \sum_{i=1}^{n} (x_a - x_p)^2$$

$$SST = \sum_{i=1}^{n} (x_a - \overline{x_a})^2$$

where x_a is the actual value of the i_{th} object, x_p is the predicted value of the i_{th} object and $\overline{x_a}$ is the actual mean value the object for n samples.

- P value (p): P value is the probability value that determines the significance of the results in relation to the null hypothesis. Since the data is not normally distributed in our case, the Levene's test calculates the p value to assess the equality of the variance for the ripeness value from different groups. It evaluates the null hypothesis that the variances in the population are equal. parities in sample variances are unlikely to have occurred based on random sampling from a population with equal variances if the resulting p value of Levene's test is less than some significance level (usually 0.05). As a result, it is determined that there is a difference between the variances in the population and that the null hypothesis of equal variances is false (Goyal and Goyal, 2012).

2.7.3 Machine Learning Approaches Used for Shelf-Life Prediction

Traditional methodologies for evaluating and forecasting the shelf life of perishable food items have long relied on static kinetic modelling, accelerated destructive testing, and periodic microbiological or chemical assays (Elizabeth *et al*., 2026). While these traditional frameworks offer scientific validity under strictly regulated laboratory environments, they frequently display significant limitations when confronted with real-world distribution networks. Storage parameters such as ambient temperature spikes, shifting relative humidity, and dynamic gas composition variations within packaged matrices can introduce non-linear degradation profiles that static mathematical formulas fail to accurately compute (Li *et al.*, 2025).

To overcome these barriers, contemporary food science has increasingly integrated data-driven Machine Learning (ML) techniques. By utilizing algorithms that excel at high-dimensional pattern recognition, researchers can seamlessly map multi-parameter environmental inputs to complex biochemical degradation indicators. This section reviews the fundamental machine learning architectures deployed within modern food quality monitoring and discusses their applicability to

Indian confectionery products like *Kaju Katli*, which are highly susceptible to moisture variations, microbial degradation, and lipid oxidation.

Artificial Neural Networks (ANN) in Kinetic Modelling

Artificial Neural Networks (ANNs) represent one of the most widely implemented non-parametric frameworks utilized in modern predictive food microbiology and shelf-life analysis (Li *et al.*, 2024). Inspired by biological neural processing, ANNs are uniquely engineered to simulate complex, non-linear relationships without requiring prior structural assumptions regarding the underlying kinetic mechanisms.

- **Handling Variable Storage Conditions:** In typical shelf-life scenarios, sensory scores, peroxide values, and free fatty acid concentrations deteriorate in a non-monotonic fashion. ANNs utilize interconnected layers of neurons (input, hidden, and output layers) equipped with non-linear activation functions (such as Sigmoid or Rectified Linear Units) to iteratively minimize error profiles between experimental datasets and predicted attributes.
- **Integration with Instrumental Analysis:** Researchers routinely feed physicochemical inputs-such as moisture content, water activity (a_w), storage temperature, and packaging barrier properties-into an ANN framework to output a continuous numerical estimation of remaining shelf life.

  The primary advantage of ANNs lies in their adaptability; as new experimental or ambient logging data becomes available, the network can be continually optimized to improve generalization capabilities across diverse and fluctuating supply chains (Li *et al.*, 2024).

Tree-Based Ensemble Methods: Random Forest and Gradient Boosting

While deep neural architectures require massive computational datasets to avoid overfitting, tree-based ensemble methods provide highly robust alternative approaches that deliver exceptional predictive performance on medium-scale experimental food science matrices.

Random Forest (RF) Regressors

Random Forest algorithms operate by constructing a multitude of independent decision trees during the training phase and outputting the mean or modal prediction of individual estimators. This ensemble approach dramatically dampens the variance and susceptibility to noise that often plagues solo decision trees.

- **Feature Importance Analysis:** In modern food preservation pipelines, RF models have demonstrated up to 98% predictive accuracy when synthesizing multi-faceted features such

as packaging configurations, gas barriers, initial preservative concentrations, and thermal controls (Li *et al.*, 2024).

- **Interpretable Metrics:** Through the integration of game-theoretic explanatory tools like SHAP (SHapley Additive exPlanations), tree-based models don't just output an expiry date; they provide clear insights into which environmental vectors (e.g., elevated humidity versus oxygen permeability) are driving product degradation at any given milestone (Li *et al.*, 2024).

Gradient Boosted Decision Trees (GBDT)

Unlike RF, which builds trees in parallel, Gradient Boosting frameworks (such as XGBoost or LightGBM) construct decision trees sequentially. Each subsequent tree is specifically trained to minimize the residual errors generated by its predecessor. This iterative optimization makes GBDTs incredibly powerful at capturing sharp threshold transitions, such as the exact point where lipid oxidation triggers rancid off-flavors in fat-rich confections.

Deep Learning and Sensor Fusion Frameworks

With the rapid proliferation of Internet of Things (IoT) hardware and non-destructive testing technologies, food quality monitoring has advanced from offline laboratory analysis to real-time, edge-computed quality forecasting.

- **Convolutional Neural Networks (CNNs) for Visual Inferences:** Deep learning models, specifically CNNs, are frequently deployed alongside optical imaging systems to capture early, non-destructive visual indicators of spoilage. For instance, integrated CNN frameworks have achieved classification accuracies as high as 95% when evaluating surface texture alterations, color degradation, or early fungal colonization on perishable food surfaces (Sonwani *et al.*, 2022).
- **Multichannel Electronic Noses (E-Noses):** Chemical spoilage often precedes visible degradation. Modern intelligent supply setups utilize multichannel gas sensors capable of tracking specific volatile organic compound (VOC) fingerprints, such as methane, carbon monoxide, or nitrogen dioxide, alongside physical variables like temperature and humidity.
- **Edge AI Execution:** By compiling lightweight machine learning frameworks (TinyML) directly onto microcontrollers, real-time sensor streams can undergo immediate local inference. This eliminates the latencies associated with continuous cloud transmission, providing logistics operators with instantaneous updates on product freshness scores.

2.7.4 Previous Studies on Machine Learning-Based Shelf-Life Prediction

In recent years, the integration of Machine Learning (ML) algorithms has revolutionized food science by replacing traditional, time-consuming mathematical degradation models-such as the Arrhenius equation and zero- or first-order kinetic equations-with dynamic, data-driven frameworks. Food spoilage is inherently complex and non-linear, often governed by simultaneous chemical, microbial, and environmental interactions. Traditional statistical methods struggle to account for all of these fluctuating variables concurrently. To address this limitation, researchers have increasingly deployed ML architectures to recognize complex patterns, map multidimensional data, and forecast food degradation with high precision. By leveraging algorithms ranging from classical statistical regressors to advanced deep neural networks, recent studies have successfully transformed shelf-life prediction from a static laboratory analysis into an automated, predictive tool capable of real-time quality tracking.

Across contemporary literature, research articles generally present previous studies on food shelf-life prediction through two distinct lenses: experimental parameter tracking and algorithmic benchmarking. Early publications in this domain focus heavily on food matrix characterization, presenting shelf-life data through laboratory-measured degradation curves of specific chemical, physical, or microbial attributes over time. However, modern machine learning-centric articles structure their literature reviews by comparing the predictive performance of different mathematical models, using evaluation metrics such as the Coefficient of Determination (R^2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). Furthermore, a significant trend in recent studies is the categorization of literature based on the non-destructive nature of the input features-distinguishing between models that require invasive chemical testing and those that leverage real-time, non-invasive environmental data like ambient temperature, gas concentrations, and relative humidity.

CNN-based ripeness classification model for Nendran Banana

Arunima *et al*., (2024) developed a custom nine-layer Convolutional Neural Network (CNN) model explicitly designed for the automated post-harvest grading and ripeness classification of the *Musa* (cv. Nendran) banana. Monitoring bananas over a 14-day shelf-life period under controlled environmental conditions (28 ± 2) degree Celsius and 85-90% RH), the researchers curated a comprehensive dataset of 4,320 images. The dataset was classified into four distinct maturity phases: unripe, medium ripe, ripe, and overripe. To maximize classification metrics, hyperparameter tuning was heavily utilized, identifying an optimal configuration consisting of a batch size of 32, a learning rate of 0.002, the Adam optimizer, and a 0.2 dropout rate.

The custom CNN model yielded an outstanding overall testing accuracy of 95%,

demonstrating robust class-wise performance with F1-scores stretching between 94% and 96%. Furthermore, the proposed architecture significantly outperformed established transfer learning models, including VGG16 (86%), VGG19 (83%), InceptionV3 (84%), ResNet50 (23%), and EfficientNetB0 (23%). The study concludes that specialized deep learning models provide highly effective, non-destructive sorting solutions capable of mitigating post-harvest quality losses.

Quality and shelf-life prediction of cauliflower using machine learning (ANNs)

Hosen and Galib (2025) explored the use of machine learning to predict the post-packaging quality and shelf-life of highly perishable cauliflower using cost-effective vacuum and chemical-based modified atmosphere packaging (MAP) techniques. Fresh cauliflower samples were monitored over a seven-day period, recording baseline and post-storage values for total soluble solids (TSS), pH, weight loss, and visual color transformation. Using analysis of variance (ANOVA) for feature selection, the authors identified TSS, weight loss, and color change as statistically significant predictors, whereas pH was excluded due to minimal variance ($p = 0.349$). To overcome visual profiling noise caused by empty gaps between florets, a custom image-processing framework was introduced, pairing a noise-reducing bilateral filter with Particle Swarm Optimization (PSO) and Markov Random Field (MRF) segmentation. This system achieved an outstanding Intersection over Union (IoU) score of 0.96. Predictive modelling comparing linear regression and Artificial Neural Networks (ANN) demonstrated the superiority of the ANN models. The network effectively captured non-linear dependencies to establish highly precise degradation curves, yielding robust R-squared values of 0.952 for TSS, 0.992 for weight loss, and 0.981 for color change relative to safe marketing thresholds

Real-time fruit freshness and quality grading with the help of AI-based framework

(Bhargava and Bansal, 2021) proposed an artificial intelligence-driven framework for real-time fruit freshness and quality grading, evaluated using an apple dataset consisting of pristine and damaged samples. The study benchmarked conventional machine learning algorithms against deep learning architectures. Within the traditional machine learning pipeline, RGB images were converted to grayscale, and key descriptors-including pixel mean intensity, color moments, and Histogram of Oriented Gradients (HOG)-were extracted and optimized using Principal Component Analysis (PCA). Performance evaluations revealed that K-Nearest Neighbors (KNN) consistently outperformed Support Vector Machines (SVM), with standalone HOG features paired with KNN achieving the highest traditional classification accuracy of 88%. To maximize sorting throughput, a Sequential Convolutional Neural Network (CNN) was developed, featuring three convolutional

layers (64, 128, and 256 nodes), max pooling, and four dense layers optimized via Stochastic Gradient Descent (SGD). The custom CNN model achieved a superior validation accuracy of 97.57% within 20 epochs. Additionally, the architecture integrated real-time text-to-speech audio alerts indicating the product state, showing immense potential for scalable deployment in industrial retail sorting lines.

Shelf-life prediction for Indian mango roll using accelerated shelf-life testing (ASLT) shelf-life plot approach

Jimeno *et al*., (2020) investigated the shelf-life prediction of Indian mango rolls using Accelerated Shelf-Life Testing (ASLT) combined with a predictive mathematical shelf-life plot approach. The confectionery fruit rolls were packed in protective aluminium foil with a polyethylene liner and stored across three temperature profiles: 30°C (ambient target), 35°C, and 45°C. Degradation kinetic matrices monitored microbiological counts, physicochemical changes (moisture content and water activity), and sensory scores evaluating flavor, color, and overall acceptability. While all experimental batches maintained optimal microbiological stability owing to the low water activity characteristic of intermediate-moisture foods (a_w=0.617), terminal shelf-life was dictated strictly by sensory panel rejection caused by severe high-temperature carotenoid browning and extreme flavor souring. The empirical trial established actual storage limits of 69 days at 45°C, 244 days at 35°C, and 255 days at 30°C. Conversely, a predictive linear regression model mapping the logarithmic shelf-life plot projected a theoretical expectancy of 462 days at 30°C. This distinct overestimation highlights the vital requirement to validate predictive mathematical equations against volatile real-world organoleptic fluctuations.

Machine Learning for Predicting Shelf Life of Burfi

Goyal and Goyal (2012) investigated a machine learning approach to automate and streamline shelf-life estimation for *burfi*, a popular traditional Indian milk confection. Recognizing that manual sensory evaluation is often subjective and resource-intensive, the authors developed recurrent artificial neural network models to track product degradation during storage at 30°C. The predictive framework utilized five key chemical and physicochemical markers as inputs-moisture, titratable acidity, free fatty acids, tyrosine, and peroxide value-to forecast the final overall acceptability score of the confection. By effectively processing these non-linear multivariate inputs, the trained models achieved exceptional precision, demonstrating a near-perfect correlation between the actual and predicted sensory scores. The study concluded that deploying automated neural networks provides a rapid, objective, and low-cost alternative to traditional laboratory testing, allowing dairy processors to reliably monitor quality retention and guarantee food safety before the

product reaches consumers.

Deep CNN Model for Predicting shelf-life of Fresh Fruits and Vegetables Using Temperature Simulation data

Baskar *et al*., (2024) explored the implementation of a deep learning-based framework to address post-harvest storage and logistical inefficiencies, focusing on real-time shelf-life prediction for fresh fruits and vegetables. Moving away from static, traditional storage guidelines that regularly result in significant food spoilage and financial losses, this approach leveraged deep Convolutional Neural Networks (CNNs) trained explicitly on simulated time-series temperature data collected during transport and warehouse phases. The predictive model utilizes statistical preprocessing alongside Analysis of Variance (ANOVA) to actively assess and evaluate how temperature fluctuations impact remaining product lifespan across differing distribution routes. By dynamically extracting patterns from live thermal streams, the system achieves highly adaptive, real-time tracking capabilities. This architecture serves as a data-driven soft-sensor tool designed to trigger instant route-optimization alerts and actionable recommendations for dynamic climate adjustments. Ultimately, the integration of real-time CNN predictions facilitates collaborative logistics management across the food supply chain, helping stakeholders minimize economic waste and safely extend commercial produce distribution boundaries.

2.8. IMPORTANCE OF SELECTED VARIABLES IN SHELF-LIFE PREDICTION

Developing a robust and intelligent machine learning model for predicting the shelf-life of intermediate-moisture dairy confections requires a carefully curated dataset that reflects the multi-dimensional nature of food spoilage. Traditional, singular evaluation methods often fail to capture the complex, concurrent changes that occur within a food matrix during storage. To overcome this limitation and ensure high predictive accuracy within our ensemble random forest framework, a targeted combination of physicochemical and microbiological variables was established. The selected feature set includes water activity (a_w), moisture content, free fatty acids (FFA), standard plate count (SPC), and yeast and mold count (YMC). Each of these parameters acts as a critical indicator of a specific degradation pathway, ranging from structural staling and chemical rancidity to biological contamination. By feeding these interconnected variables into the random forest ensemble, the algorithm can effectively map the non-linear kinetics of product decay and accurately isolate how individual environmental stresses accelerate quality loss. The following subsections directly justify the selection of these specific parameters and detail their fundamental importance in predicting the consumer acceptability and safety threshold of kaju katli.

- **Water Activity (a_w)**

Water activity (a_w) was selected as a critical input feature for our ensemble random forest model because it measures the energy state of water within the food matrix rather than the total amount of water present. In a rich, intermediate-moisture dairy confection like *kaju katli*, water activity serves as the primary thermodynamic gatekeeper for quality degradation. It directly controls the minimum thresholds required for microbial survival, enzymatic reactivity, and chemical breakdown. By tracking a_w, the ensemble model can accurately capture non-linear degradation tipping points, such as the exact threshold where pathogenic or spoilage organisms become active, or where structural staling begins. Including this parameter provides the machine learning algorithm with a highly sensitive indicator of environmental vulnerability, allowing it to predict how slight shifts in storage humidity will accelerate shelf-life termination long before visible spoilage appears.

- **Moisture Content**

While water activity dictates chemical viability, moisture content represents the total quantitative mass of water held within the *kaju katli*. We selected this parameter because it is the fundamental driver behind the confection's coveted texture, directly affecting characteristics like softness, chewiness, and mouthfeel. Over time, moisture migration between the product and its packaging causes distinct quality defects, ranging from surface weeping to severe moisture loss that yields a dry, crumbly texture. Furthermore, moisture content exhibits complex, non-linear interactions with raw ingredient variables. By integrating total moisture into our random forest ensemble, the model can effectively cross-reference mass-based water loss against structural stability. This enables the algorithm to isolate whether a drop in quality scores is caused by physical desiccation or separate chemical changes, ensuring highly accurate texture-based shelf-life boundaries.

- **Free Fatty Acids (FFA)**

Because cashew nuts and ghee contain a high concentration of lipids, *kaju katli* is highly susceptible to hydrolytic and oxidative rancidity. Free Fatty Acids (FFA) were chosen for our predictive model as the primary chemical indicator of lipid oxidation and fat degradation. As triglycerides break down due to moisture, heat, or lipolytic enzymes, FFA levels rise, generating volatile compounds responsible for off-flavors, pungent odours, and a distinctive soapy taste. This chemical shift completely destroys consumer acceptability well before any physical or microbial changes occur. The random forest ensemble leverages FFA data to map the subtle, early-stage chemical kinetics of rancidity. This allows the model to capture the progression of lipid oxidation over time and accurately forecast when the confection will become unpalatable, establishing a robust chemical threshold for shelf-life prediction. A direct indicator of lipolytic and oxidative rancidity in high-fat nut products. AOAC (2005) and FSSAI

Manual (2016) recommend FFA as the primary chemical index for nut-based confectionery shelf life. Lima et al. (2012) correlated FFA increase with sensory rejection in cashew sweets.

- **Standard Plate Count (SPC)**

Standard Plate Count (SPC) was selected to serve as the baseline biological index within our machine learning framework, indicating the total viable aerobic microbial load present on the confection. Throughout storage, the blending of milk solids, processed cashews, and sugar creates a nutrient-rich environment where environmental contaminants can multiply if storage temperatures fluctuate. A rising SPC directly correlates with general quality loss, batch-to-unit cross-contamination, and structural breakdown. By feeding SPC trends into our ensemble random forest model, the algorithm gains a direct window into the overall hygienic status and progressive biological decay of the product. This enables the model to effectively weigh microbial growth rates against physicochemical variables, guaranteeing that the final shelf-life estimation reflects strict compliance with public health and food safety standards.

- **Yeast and Mold Count (YMC)**

Yeast and Mold Count (YMC) was chosen as an essential microbiological feature because fungal spoilage is the most common visual defect that terminates the shelf-life of intermediate-moisture sweet confections. Due to the high osmotic pressure exerted by sugar and the moderate water activity within *kaju katli*, typical bacteria are often suppressed, leaving the product highly vulnerable to xerophilic molds and osmophilic yeasts. Fungal contamination manifests as visible surface colonies, structural softening, and localized fermentation defects that render the product entirely unmarketable. Incorporating YMC into the ensemble random forest model provides the algorithm with a targeted biological warning system. The random forest blocks can trace the exponential growth curves unique to fungi under varying storage conditions, ensuring the model accurately identifies the precise day a batch faces aesthetic or biological rejection. FSSAI and IS 5402/IS 5403 mandate microbial limits for milk- and nut-based sweets. Kumar *et al*., (2017) and Jay *et al*., (2005) identified microbial count as the limiting factor for shelf life of khoa- and nut-based sweets

These variables were selected because they collectively represent the physicochemical and microbiological quality changes occurring during storage. By combining thermodynamics, mass variables, lipid chemistry, and targeted microbial tracking, this multi-dimensional feature set provides our ensemble random forest model with the comprehensive data required to deliver highly accurate, robust, and dependable shelf-life predictions for kaju katli.

2.9 RANDOM FOREST ALGORITHM: ARCHITECTURE AND METHODOLOGICAL JUSTIFICATION

2.9.1 Definition and Core Principles

The Random Forest algorithm is a robust machine learning technique that belongs to the broader category of ensemble learning methods. Rather than relying on the predictive performance of a single, isolated statistical model, ensemble methods combine the predictions of multiple individual models to generate a more accurate, stable, and generalized output. Specifically, a Random Forest operates as a collective ecosystem composed of numerous distinct decision trees. Each individual decision tree within the forest acts as a base classifier or regressor that splits data into hierarchical branches based on feature thresholds. By constructing an extensive collection of these deeply uncorrelated decision trees, the algorithm effectively leverages the "wisdom of the crowd." The fundamental principle underlying this architecture is that while individual trees may produce localized errors or exhibit high variance, the combined ensemble dramatically diminishes these errors, smoothing out statistical anomalies to deliver highly dependable predictions.

2.9.2 Operational Mechanics and Training Workflow

The training mechanism of a Random Forest relies on two primary data-level and structural operations: bootstrap aggregating (bagging) and random feature selection. The process unfolds systematically across the following computational phases:

1. **Bootstrap Sampling:** Given an original training dataset, the algorithm creates multiple sub-datasets through a process called bootstrapping. This involves sampling observations uniformly and with replacement. Consequently, each generated subset is the same size as the original data but contains duplicate rows, while roughly one-third of the original observations are left out-termed Out-of-Bag (OOB) data. Each individual decision tree in the forest is trained exclusively on its respective bootstrap sample.
2. **Random Feature Subspace Selection:** Unlike standard decision trees that evaluate every available feature to determine the optimal branch split, Random Forest injects an extra layer of randomness. At each node split during a tree's growth, the algorithm restricts the search space to a randomly selected subset of features. This intentional constraint ensures that the trees become highly uncorrelated with one another, preventing a single dominant feature from dictating the structure of every tree in the forest.
3. **Ensemble Aggregation (Average Prediction):** Once all individual decision trees are fully grown without pruning, the forest is ready to make predictions on unseen testing data. When a new sample input is fed into the model, it passes down every decision tree simultaneously. In a regression task-such as forecasting numerical shelf-life boundaries-each tree outputs a specific numerical value. The final output of the Random Forest is

computed by taking the mathematical average of all predictions generated across the entire collection of trees, effectively cancelling out localized variance.

2.9.3 Key Advantages in Predictive Analytics

The Random Forest algorithm offers distinct mathematical and computational benefits over linear architectures, making it a highly favoured tool in food science and predictive modelling:

- **Handling of Complex Nonlinear Relationships:** Traditional linear regression assumes straight-line dependencies between parameters, which rarely matches reality in biological systems. Random Forest naturally maps multi-dimensional, step-like, and highly nonlinear degradation pathways without requiring manual mathematical transformations.
- **Proficiency with Small Datasets:** While deep learning models require tens of thousands of data points to achieve convergence, Random Forest can build stable decision boundaries on remarkably small datasets with a limited number of total observations.
- **Robustness to Structural Noise:** Real-world sensory and laboratory datasets often contain statistical outliers, experimental anomalies, or reading noise. Because the final prediction is an average of hundreds of trees, isolated errors within a few branches have an extremely negligible impact on the collective output.
- **Mitigation of Overfitting Tendencies:** Deep, unpruned single decision trees are notorious for memorizing training data noise (overfitting). By combining bagging with random feature selection, Random Forest creates a structural ceiling that prevents the variance of individual trees from degrading its generalization capabilities on new validation data.
- **Automated Feature Importance Analysis:** The algorithm provides an internal validation mechanism by tracking how much the prediction error increases when a specific variable's values are permuted or when a node split is made. This allows researchers to mathematically rank which physicochemical or biological factors exert the strongest influence on shelf-life degradation.

2.9.4 Justification for Selection in the Present Study

Random Forest was selected because the available dataset contained a limited number of observations and multiple interacting physicochemical and microbiological variables. In a delicate confection like *kaju katli*, quality degradation is not driven by a singular, isolated parameter; rather, it is dictated by the intricate, simultaneous interplay between water activity shifts, moisture migration, lipid oxidation (FFA accumulation), and microbial kinetics (SPC and YMC growth). The

algorithm can model nonlinear relationships, is less susceptible to overfitting, and performs effectively on small datasets, making it exceptionally suitable for the shelf-life prediction of *kaju katli*. By choosing this ensemble framework, the study guarantees a highly reliable predictive tool capable of handling multivariate biological matrices without suffering from the data-scarcity constraints common in specialized shelf-life trials.

2.10 RESEARCH GAP IDENTIFIED

A comprehensive evaluation of the existing literature reveals a distinct imbalance between empirical food preservation studies and the application of advanced predictive computer science models within the ethnic dairy and confectionery sector. While the application of machine learning (ML) architectures has progressively expanded into mainstream food analytics, traditional Indian confections (mithai) remain remarkably underrepresented in the domain of automated shelf-life forecasting.

A foundational step in this domain was taken by Goyal and Goyal (2012), who successfully developed a specialized machine learning framework using the Elman recurrent neural network technique to predict the shelf-life of burfi stored at 30°C. Their predictive model processed specific chemical and physical inputs-such as moisture content, titratable acidity, free fatty acids, tyrosine, and peroxide value-to accurately estimate the confection's overall sensory acceptability score. Beyond burfi, separate neural network frameworks have been successfully applied to alternative dessert systems, such as cakes, kalakand, and pistachio-jewelled desserts. This establishes a clear precedent that artificial intelligence can effectively replace expensive, time-consuming, and cumbersome laboratory testing for traditional sweets.

However, a critical research gap emerges when shifting focus from standard milk-solid confections to premium, high-lipid nut pastes-specifically kaju katli. Although generalized post-harvest storage studies and traditional laboratory-based shelf-life evaluations have been conducted historically to observe the physical staling and rancidity kinetics of cashew confections, a dedicated machine learning predictive model tailored specifically for kaju katli has not yet been developed. This oversight represents a major technological limitation due to the structural and chemical differences between these products:

- **Distinct Formulation Matrix:** Unlike standard plain burfi, which primarily blends buffalo milk solids and sucrose, kaju katli consists of a high-fat cashew paste core.
- **Unique Chemical Vulnerabilities:** The elevated concentrations of lipids introduced by cashew nuts make kaju katli far more susceptible to rapid oxidative and hydrolytic rancidity than standard dairy products.

- **Complex Spoiling Interplay:** Fungal spoilage patterns inside the intermediate-moisture environment of nut confections differ fundamentally from the degradation curves seen in milk-solid confections.

Consequently, simply extrapolating existing burfi or milk-cake neural networks to kaju katli yields highly inaccurate predictions.

Furthermore, the existing literature on mithai shelf-life modelling relies almost exclusively on single or multilayer feedforward and recurrent neural networks trained via Bayesian regularization. While effective, these studies routinely omit modern ensemble-based machine learning approaches, such as Random Forest. This architectural limitation leaves another noticeable gap: existing models lack automated feature importance metrics capable of ranking which exact parameter (e.g., lipid oxidation versus microbial load) accelerates degradation fastest under localized environmental stress. By identifying these deficiencies, this study establishes its necessity. Developing an ensemble Random Forest model explicitly optimized for kaju katli bridges the divide between classical food science storage studies and modern predictive analytics, offering the dairy and sweet processing industry a tailored, highly precise tool for real-time quality assurance.

CHAPTER III
MATERIALS AND METHODS

This chapter covers the preparation method adopted for preparing kaju katli sample for data collection, tools and methodology used for collecting data related to the shelf-life study and data collection, feature engineering used. In the present study, Kaju Katli samples were prepared using a standardized formulation to ensure consistency in quality and facilitate the development of a machine learning-based shelf-life prediction model. A standard recipe containing 65% cashew nut powder and 35% sugar was followed throughout the study. The prepared samples were subsequently subjected to storage studies under different environmental conditions, and their physicochemical and sensory characteristics were monitored for model development (Sharma, 2017).

3.1. MATERIALS AND SAMPLE PREPARATION

Cashew nut pieces, off white in color, without stalks, free from rancid off flavour) and commercially available white crystalline sugar, free from dirt, were obtained from a grocery store near KCAEFT. Soaked cashew splits 500 g was ground for 1 min using a laboratory mixture grinder (Make: Sujata Dynamix 900 Watt) before 28% potable water addition. The sample was then further grounded for 12 min to a uniform particle size. After that the cashewnut paste was transferred into non-stick Teflon coated plan. Sugar was added and mixed manually with help of a hand scraper while heating using a gas stove for desired temperature of 100 °C. The mixture was stirred manually with help of scraper until it starts leaving the surface of the vessel. The cooked mixture was then cooled in another vessel up to 45 °C with continuous stirring for about 7-8 min. The tempered mass was then rolled using hand roller (bellan) to the thickness of 5mm and cut by knife in the diamond shape having size of 3×3 cm. Preparation steps with fixed parameters are shown in fig 3.1

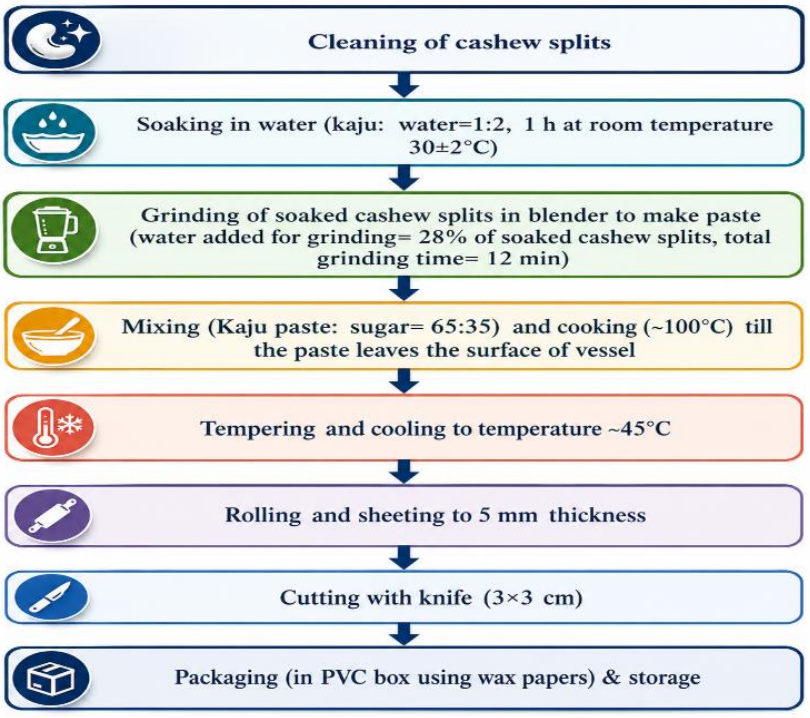


Fig. 3.1. Steps in preparation of kaju katli sample

3.2. SHELF-LIFE CHARACTERSITICS ANALYSIS OF KAJU KATLI SAMPLES

The study commenced with a systematic review of published literature to identify the critical quality indicators governing the shelf life of traditional Indian sweets, particularly sugar-fat confections such as Kaju Katli (cashew fudge). Research articles, standard guidelines from the Food Safety and Standards Authority of India (FSSAI), Bureau of Indian Standards (BIS) specifications, and peer-reviewed journals on confectionery science and food microbiology were surveyed. The review encompassed studies on lipid oxidation, moisture-mediated deterioration, microbial spoilage, and water activity relationships in nut-based and sugar-based food matrices.

Based on the reviewed literature, the following five physicochemical and microbiological parameters were identified as the most relevant predictors of quality deterioration and shelf life in Kaju Katli:

- Free Fatty Acid (FFA) content - a key chemical indicator of lipolytic rancidity arising from hydrolysis of cashew lipids, commonly reported as percent oleic acid.
- Moisture Content (MC) - governs texture, stickiness, and susceptibility to microbial growth; particularly critical in sugar-fat confections prone to moisture migration.

- Water Activity (a_w) - the thermodynamic availability of water determining microbial growth potential and chemical reaction rates, independent of total moisture content.
- Standard Plate Count (SPC) - a quantitative measure of total viable aerobic bacteria, indicative of overall hygienic quality and product safety.
- Yeast and Mould Count (YMC) - a fungal contamination indicator particularly relevant for sugar-rich matrices susceptible to osmophilic and xerophilic organisms.

This multi-parameter approach was adopted to capture both the chemical and microbiological dimensions of spoilage, enabling a more robust and comprehensive shelf-life prediction model. (Ankita *et al*., 2016).

Kaju Katli samples were prepared with a standardised formulation consisting of 65% cashew nut paste and 35% sucrose by weight, without addition of any preservatives, artificial flavours, or colouring agents, to represent a commercially typical unpackaged product. Samples were prepared in a single batch under controlled conditions to eliminate batch-to-batch variability. Following preparation, samples were portioned uniformly and assigned to two distinct storage conditions to simulate realistic post-production environments.

Storage Conditions

To assess the influence of temperature on the rate of quality deterioration, two storage conditions were established:

- Ambient Storage: Samples were held at 32 ± 2°C in a temperature-controlled chamber replicating typical room-temperature retail and household conditions prevalent in tropical climates. Sampling was conducted over a total study period of 12 days.
- Refrigerated Storage: Samples were stored at 7 ± 2°C in a laboratory refrigerator. Sampling was conducted over a total study period of 20 days.

All samples were stored in identical food-grade containers with standardised headspace to minimise variation in atmospheric exposure between replicates.

The physicochemical properties of the kaju katli were analyzed. Parameters like moisture content, water activities, free fatty acids, texture, sugar content, and color were measured using standard methods, as described in the following sections. Each quality parameter was monitored at different intervals based on its expected rate of change and analytical sensitivity, as summarised in Table 3.1. The staggered sampling strategy was designed to maximise the information density of the dataset while managing analytical workload.

Table 3.1: Sampling schedule for each quality parameter under both storage conditions

Parameter	Sampling Frequency	Ambient (12 days)	Refrigerated (20 days)
Water Activity (aw)	Daily	Days 0-12 (daily)	Days 0–20 (daily)
Moisture Content (MC)	Every 3 days	Days 0, 3, 6, 9, 12	Days 0, 3, 6, 9, 12, 15, 18
FFA Content	Every 3 days	Days 0, 3, 6, 9, 12	Days 0, 3, 6, 9, 12, 15, 18
Standard Plate Count (SPC)	Every 3 days	Days 0, 3, 6, 9	Days 0, 3, 6, 9, 12, 15, 18
Yeast & Mould Count (YMC)	Every 4–5 days	Days 0, 4–5, 9–10, 12	Days 0, 4–5, 9–10, 14–15, 19–20

3.2.1. Determination of Moisture Content

Moisture content in Kaju Katli typically ranges between 8-15%, depending on the preparation method and storage conditions. The moisture content of *Kaju Katli* is a critical primary determinant of its textural profile, water activity (a_w), and susceptibility to microbial spoilage. Historically, the standard benchmark for determining moisture in confectionery and nut-based products is the AOAC Official Method 925.09 (AOAC, 1995), which relies on conventional hot-air or vacuum oven drying until a constant weight is achieved.

While the AOAC gravimetric method remains highly accurate, its operational timeline requires prolonged drying cycles (typically 16 to 24 hours). This creates a significant bottleneck when real-time, iterative data is required during active processing trials.

To optimize analytical throughput without sacrificing precision, this study utilized an advanced thermogravimetric approach via an Infrared Moisture Analyzer (Shimadzu MOC63u Moisture Analyzer). This instrument combines infrared heating and an ultra-precise internal analytical balance to continuously monitor weight loss in real-time, reducing analytical processing time from hours to minutes.

Plate 3.1 Infrared moisture analyzer

Instrument Calibration and Standard Operating Procedure (SOP)

To ensure the integrity, repeatability, and traceability of the experimental data, a strict calibration protocol and standard operating procedure were executed prior to sample analysis.

Instrument Calibration

- **Weight Calibration:** The internal electronic balance of the Infrared Moisture Analyzer was calibrated using a certified standard 50 g non-magnetic stainless-steel check-weight. The calibration was accepted only when the instrument readout matched the standard mass precisely within a tolerance limit of ±0.001g.
- **Temperature Verification:** The infrared heating module was verified using a specialized temperature calibration kit to ensure that the actual operational temperature inside the heating chamber closely mirrored the digital setpoint ±1°C.
- **Baseline Taring:** Before every individual run, the empty aluminium sample pan was placed in the chamber, heated to burn off residual moisture, cooled, and tared to zero out any baseline drift.

General Standard Operating Procedure (SOP)

1. The instrument was switched on and allowed a warm-up period of 15 minutes to stabilize the internal electronic components and infrared source.
2. The operational parameters-including heating profile (Standard Mode), target temperature (105°C), and shut-off criteria (Automatic slope detection)-were programmed into the micro-controller interface.
3. The protective wind shield was kept closed during all idle and measurement phases to eliminate errors caused by ambient laboratory air currents.

Method for Infrared Moisture Analysis (Shimadzu MOC63u Moisture Analyzer)

The analytical protocol for measuring the moisture content of the Kaju Katli samples using the Infrared Moisture Analyzer was standardized as follows:

- **Sample Preparation:** Due to the composite nature of Kaju Katli (containing cashews, sugar matrix, and ghee fats), achieving a uniform surface area is vital. A representative portion of the sample was carefully grated and homogenized using a sterile mortar and pestle immediately before testing to prevent premature moisture loss to the ambient environment.
- **Sample Loading:** Approximately (3.00 ± 0.50 g) of the homogenized sample was evenly spread in a thin, uniform layer across the surface of the pre-tared aluminium pan. Thick clumps were avoided to prevent surface charring and incomplete internal drying.
- **Analysis Phase:** The sample pan was placed back onto the instrument's weighing core, and the chamber lid was securely closed to initiate the automatic test cycle. The infrared lamps subjected the sample to a steady temperature of (105°C), vaporizing the free and loosely bound water within the product matrix.
- **Endpoint Determination:** The instrument tracked the weight loss curve continuously. The analysis was governed by an automatic switch-off criterion (constant weight logic): the heating cycle automatically terminated when the continuous mass loss dropped below 1 mg over a rolling window of 60 seconds.
- **Data Processing:** Upon completion, the integrated microprocessor calculated the total moisture loss based on the initial and final mass values. The final value was recorded directly from the digital display interface as a percentage of total moisture content on a wet-basis (wb).

The fundamental calculation computed automatically by the MCA100-IR system follows the formula:

$$Moisture\ Content\ (\%,\ wb) = \left(\frac{M_{initial} - M_{final}}{M_{initial}}\right) \times 100$$

Where:

- $M_{initial}$= Initial weight of the sample (g) before infrared heating.
- M_{final} = Final constant dry weight of the sample (g) at test termination.

All samples were evaluated in triplicate (n = 3) to compute standard deviation and ensure statistical reliability for the final thesis presentation. Moisture content values for both the samples, stored at ambient temperature and in refrigerated conditions were collected in every 3 days.

3.2.2 Determination of Water Activity(a_w)

Principle and Regulatory Framework

Water activity (a_w) is defined as the ratio of the vapor pressure of water in a food substrate (p) to the vapor pressure of pure water (p_0) at an identical temperature. It serves as a vital thermodynamic metric dictating microbial vulnerability, chemical stability, and the ultimate shelf life of Kaju Katli.

$$a_w = \frac{p}{p_0}$$

In this study, water activity was quantified in strict accordance with ISO 18787:2017 (Food products -Determination of water activity). The measurements were executed utilizing the benchtop AquaLab Water Activity Meter (Series 3/3TE, Decagon Devices) displayed. This instrument operates via the chilled-mirror dew point technique.

Inside a sealed headspace chamber, an infrared beam detects the exact moment condensation forms on a chilled mirror, determining the dew point temperature of the air. This temperature is paired with the sample surface temperature (measured via an internal infrared thermometer) to provide a highly precise, mathematically derived water activity value (± 0.003 a_w) independent of product matrix variations.

Instrument Calibration and Verification

To maintain compliance with ISO 18787:2017 guidelines, a verification protocol was carried out before beginning each daily testing session to check the performance of the optical mirror assembly:

- **Chamber Inspection:** The sample drawer and sensor port were checked for particulate contamination, fat films, or volatiles. The mirror was cleaned using isopropyl alcohol and distilled water wipes when necessary.
- **Standard Verification Solution:** Certified, non-saturated salt verification standards (METER Group / Decagon) mimicking the target profile of confectionery products were utilized-specifically, 0.760 a_w (6.0 mol/kg NaCl) at an analysis temperature of 25±0.2°C.
- **Acceptance Criteria:** The calibration was deemed successful and stable if the instrument display yielded a value within $\pm 0.003\ a_w$ of the standard's certified nominal value. If the deviation exceeded this limit, a formal multi-point calibration linear adjustment was executed through the system firmware controls.

Standard Operating Procedure (SOP) for Kaju Katli Analysis

The execution sequence for determining the water activity of Kaju Katli samples was standardized using the following methodology:

1. Sample Preparation

Because *Kaju Katli* contains a heterogeneous fat-sugar structure, maintaining a representative surface-area-to-volume ratio inside the sample cup is essential.

- A core section of the sweet was excised, avoiding the outer edges that may have undergone premature ambient drying.
- The sample was quickly chopped or minced into small fragments using a clean spatula. This step was performed rapidly to limit exposure to ambient laboratory air, which could alter the sample's true moisture equilibrium.

2. Loading the Sample Cup

- The minced Kaju Katli was transferred into a standard, disposable AquaLab plastic sample cup.
- **Volume Control:** The cup was filled to just under half its total capacity (approximately 2-3 g). Overfilling was strictly avoided to prevent the sample matrix from touching the internal sensor head when the drawer closed.
- **Rim Hygiene Check:** The bottom and perimeter walls of the sample cup were carefully inspected. Any stray fragments or fat residues were wiped away using a lint-free tissue to ensure a tight, clean seal inside the chamber.

3. Executing the Measurement Sequence

- The instrument knob, located on the front right of the AquaLab unit, was turned to the **OPEN/LOAD** position, and the sample drawer was carefully pulled forward.
- The prepared sample cup was seated flatly within the circular drawer nest, and the drawer was pushed back into the chassis.
- The front control knob was turned to the **READ** position. This mechanical motion lowers the internal sealing O-ring onto the cup rim, creating a completely isolated micro-environment headspace.
- The activation of the sealed chamber engages the optical chilled-mirror system. The digital LCD screen monitored real-time equilibration progress.

4. Data Logging and Replication

- The sample remained undisturbed while the internal fan accelerated moisture equilibrium over the mirror.
- Once the vapor pressure stabilized, the instrument generated an audible notification signal and locked the final equilibrium water activity score on the LCD panel.
- The final a_w value and corresponding sample temperature (°C) were recorded directly into the project log.

- After recording, the front knob was rotated back to the **OPEN** position, the cup was removed and discarded, and the chamber was left sealed until the next test run.

In alignment with ISO 18787:2017 data accuracy guidelines, every batch of Kaju Katli was tested in triplicate (n=3) every day for both the samples. The average of these three runs was calculated to establish the true water activity value used in the shelf-life predictive models.

Plate 3.2 Water activity meter

3.2.3 Determination of Free Fatty Acids

Cold Solvent Isolation of Fat from Confectionery Matrix

The quantitative and qualitative analysis of lipids embedded within a composite carbohydrate-protein-fat matrix such as Kaju Katli requires an isolation method that prevents the chemical alteration of native constituent fats. Conventional hot-solvent extraction techniques (such as Soxhlet extraction involving continuous thermal cycling) subject the sample to prolonged thermal stress. In products characterized by a high reducing sugar content and unsaturated fatty acid profiles, this thermal exposure can induce artificial lipid oxidation, thermal hydrolysis, or localized sugar caramelization, thereby skewing baseline Free Fatty Acid (FFA) values.

To mitigate these analytical errors, a cold solvent extraction methodology was deployed. This protocol aligns with the fat isolation guidelines specified in IS 548 (Part 1): 1964 (Reaffirmed 2021) - Methods of Sampling and Test for Oils and Fats and constitutes an approved modification of AOAC Official Method 948.22 (Fat in Nuts and Nut Products). The use of n-hexane at ambient room temperature selectively solubilizes non-polar triacylglycerols while leaving the structural carbohydrate polymers and milk/nut proteins insoluble, preserving the chemical integrity of the native fat for downstream quality assays.

Standard Operating Procedure (SOP)

1. **Sample Maceration:** A representative 10.0 g sample of Kaju Katli was ground into a uniform, fine paste using a clean laboratory mortar and pestle to maximize the total surface area exposed to the extraction solvent.
2. **Solvent Dissolution:** The homogenized paste was quantitatively transferred into a 250 mL

glass conical flask, followed by the addition of 100 mL of analytical-grade n-hexane. The flask was sealed securely with a stopper to prevent volatile solvent loss.

3. **Mass Transfer and Agitation:** The mixture was subjected to continuous, vigorous agitation on a mechanical platform shaker for 30 minutes. This step ensured complete penetration of the n-hexane into the cashew matrix to fully solubilize the embedded lipids.
4. **Filtration:** The mixture was filtered through a dry Whatman No. 1 filter paper supported by a glass funnel into a clean, pre-weighed receiving Erlenmeyer flask. The insoluble structural residues remained on the filter bed, while the filtrate containing the liquefied lipids and n-hexane was collected.
5. **Solvent Desolventization:** The collected filtrate was placed in a digitally regulated water bath maintained strictly at 65°C (under a well-ventilated fume hood). This temperature successfully vaporized the n-hexane (boiling point: 68.7°C), leaving behind a purified, isolated lipid phase.

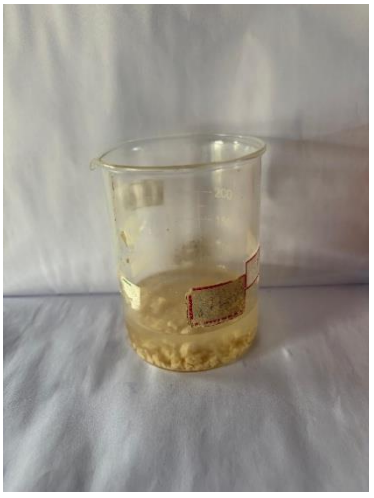


Plate 3.3 Maceration of Kaju Katli Sample

Plate 3.4 Dissolution of Kaju Katli Sample in n-Hexane

Plate 3.5 Extracted Fat Dissolved in 95% (v/v) Ethanol

3.2.4 Quantitative Estimation of Free Fatty Acids (FFA)

Free Fatty Acids (FFA) are generated via the hydrolytic cleavage of fatty acids from their respective glycerol backbones. This degradation pathway is accelerated by intrinsic moisture, elevated storage temperatures, and native lipolytic enzymes. Because the lipid profile of Kaju Katli is dominated by oleic acid derived from cashew nut fats, quantification of FFA serves as a direct chemical indicator of hydrolytic rancidity and quality loss during storage.

The analytical procedure carried out in this study is based on the classical alkali titrimetric method, executed in strict accordance with AOAC Official Method 940.28 (Fatty Acids (Free) in Crude and Refined Oils) and IS 548 (Part 1): 1964 (Section 7 -Determination of Free Fatty Acids and Acid Value).

Reagent Preparation and Neutralization Protocol

- **Titrant Preparation:** A standard solution of 0.1 N Sodium Hydroxide (NaOH) was prepared and subsequently standardized to establish its exact normality (N).
- **Solvent System Neutralization:** Commercial 95% (v/v) ethanol was utilized as the solvating medium. To eliminate potential experimental errors caused by dissolved ambient carbon dioxide or trace organic acids within the commercial reagent, a mandatory neutralization step was performed.
- **Neutralization Execution:** A 50 mL aliquot of 95% ethanol was mixed with 3 drops of 1% phenolphthalein indicator solution in a flask. This solution was titrated dropwise with 0.1 N NaOH until a faint, stable pink color persisted for at least 30 seconds, ensuring a completely neutral solvent baseline prior to sample introduction.

Standard Operating Procedure (SOP) for Titration

1. **Sample Quantitation:** The dried, isolated fat obtained from the cold solvent extraction process was accurately weighed directly into a clean 250 mL Erlenmeyer flask. The exact mass (W) was recorded using an analytical balance to a precision of ±0.001.
2. **Matrix Solubilization:** A 50 mL of the pre-neutralized 95% ethanol was added to the flask to fully dissolve the lipid polymers. The flask was swirled thoroughly until a clear, single-phase homogeneous solution was obtained.
3. **Indicator Addition:** Exactly 3 drops of 1% phenolphthalein indicator solution were added directly to the neutralized lipid-solvent matrix.
4. **Titrimetric Measurement:** The mixture was titrated against the standardized 0.1 N NaOH solution filled in a calibrated burette under continuous manual swirling to ensure uniform reagent distribution.
5. **Endpoint Recognition:** The titrant was added dropwise as the reaction approached its transition zone. The titration was terminated at the exact point where a faint pink color emerged and remained completely stable throughout the solution matrix for a minimum of 30 seconds. The final burette reading was recorded to determine the total volume (V) of titrant consumed.

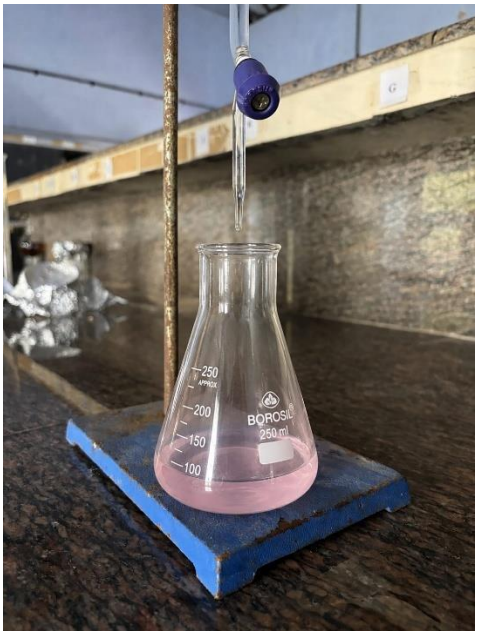


Plate 3.6 Pink colour obtained on titration of oleic acid present in sample

Mathematical Calculation

In accordance with standard regulatory codes for nut and confectionery fats, the free fatty acid content is expressed as a percentage of Oleic Acid on a mass basis (w/w). The value was computed

utilizing the following equation:

$$\text{Free Fatty Acids (as \% Oleic Acid)} = \frac{V \times N \times 28.2}{W}$$

Where:

- V = Volume of standardized NaOH titrant consumed during the assay (mL)
- N = Exact calculated normality of the NaOH titrant (nominally 0.1 N)
- 28.2 = Milli-equivalent weight of oleic acid (g/meq), derived from its molecular mass(282.47 g/mol $\div 1000 \times 100\%$)
- W = Total weight of the extracted fat sample used for the titration (g)

To ensure statistical reliability and accurate input data for the shelf-life prediction models, all extractions and titrations were performed in triplicate (n = 3) to determine the mean values and corresponding standard deviations. FFA values for both the samples, stored at ambient temperature and in refrigerated conditions were collected in every 3 days.

3.2.5. Microbiological Analysis

Enumeration of Aerobic Mesophilic Microorganisms (Standard Plate Count)

Theoretical Framework and Reference Standard

According to FSSAI (Food Safety and Standards Authority of India), Kaju katli is considered a nut-based confectionery, coming under 18th part, Indian Sweets and Indian Snacks & Savories Products, under the subcategory, Dry fruit and nut-based sweets (18.1.3) as given in Annexure 1. The Standard Plate Count (SPC) is a foundational microbiological metric used to determine the total viable aerobic bacterial load, sanitary processing quality, and progressive microbial spoilage kinetics of food matrices. Because Kaju Katli is a nutrient-dense sweet containing cashew proteins and high concentrations of simple carbohydrates, it provides an ideal substrate for the proliferation of heterotrophic mesophilic bacteria under variable supply chain environments.

In this study, the quantification of the viable aerobic microbial population was modelled around the procedural framework of ISO 4833-1:2013 (Microbiology of the food chain - Horizontal method for the enumeration of microorganisms - Part 1: Colony count at 30 °C by the pour plate technique). To accommodate the practical resource availability and operational constraints of an undergraduate technical research laboratory, specific modifications were introduced to the standard ISO protocol: Nutrient Agar was deployed as the solid growth medium instead of Plate Count Agar (PCA), and

sterile distilled water was utilized as the serial dilution vector instead of Buffered Peptone Water (BPW). These variations provided a robust analytical alternative that accurately tracked population shifts without altering the comparative integrity of the longitudinal shelf-life dataset.

Reagent Preparation and Sterilization

- **Nutrient Agar Preparation:** Dehydrated commercial Nutrient Agar medium was accurately weighed and suspended in distilled water inside an Erlenmeyer flask according to the manufacturer's structural specifications. The suspension was heated under continuous manual agitation and brought to a boil for 1 minute to ensure complete dissolution of the agar polymers.
- **Sterilization Protocol:** The flask containing the dissolved medium, along with separate glass test tubes filled with precisely measured volumes of distilled water intended for dilution blanks, was thermally sterilized via autoclaving at $121°C$ under 15 lbs for 15 minutes.
- **Thermal Stabilization:** Post-sterilization, the molten Nutrient Agar was transferred to a digitally regulated water bath maintained at $45 \pm 1°C$. This specific temperature window kept the agar fluid while preventing thermal cell death or heat shock to the bacterial cells during the subsequent pour-plating process.

Standard Operating Procedure (SOP) for Inoculation and Serial Dilution

To prevent cross-contamination from ambient laboratory air, all sample preparation, serial dilution, and plating operations were executed inside a pre-sanitized Laminar Airflow (LAF) bench under strict aseptic conditions.

1. **Primary Homogenization (10^{-1} Dilution):** A representative 10.0 g sample of *Kaju Katli* was aseptically excised using a sterile spatula and transferred into a sterile flask containing 90 mL of autoclaved distilled water. The mixture was shaken vigorously until a completely homogeneous suspension was achieved, establishing the base 10^{-1}dilution.
2. **Serial Dilution Chain:** A 1.0 mL aliquot from the primary 10^{-1} homogenate was transferred via a micropipette with a sterile tip into a test tube containing 9.0 mL of sterile distilled water to yield a 10^{-2}dilution. This process was repeated sequentially to prepare the target working dilutions of 10^{-3} and 10^{-4}.
3. **Inoculation:** Exactly 1.0 mL aliquots from the 10^{-3} and 10^{-4} dilution tubes were dispensed directly into the centres of duplicate, pre-labelled sterile glass Petri dishes.
4. **Pour Plating:** Approximately 15 to 20 mL of the stabilized, molten Nutrient Agar ($45 \pm 1°C$)

was poured over the liquid sample inoculum in each Petri dish. The plates were immediately swirled using a gentle, flat manual motion (alternating clockwise, counter-clockwise, and figure-eight paths) to blend the sample uniformly within the liquefied agar matrix before it set.

5. **Incubation:** The plates were left undisturbed on a level bench to solidify completely. Once set, the Petri dishes were inverted and placed in a microbiological incubator maintained at $30 \pm 1°C$ for an analytical window of 48 ± 2 hours.

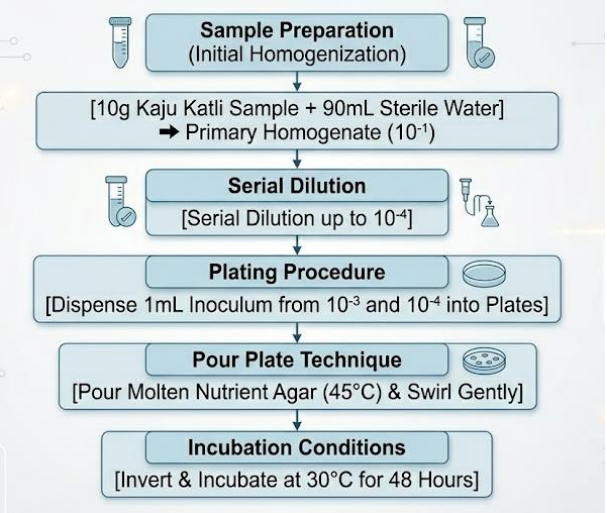


Fig. 3.2 Flowchart of process followed for TPC analysis

Colony Counting and Quantification

Following the 48-hour incubation period, the plates were removed and evaluated using a digital colony counter. In alignment with standard microbiological counting criteria, only plates containing between 30 and 300 distinct colonies were selected for final quantification. The concentration of viable aerobic mesophilic bacteria per unit mass of the sweet was computed using the following equation:

$$\text{Colony Forming Units (CFU/g)} = \frac{N}{V \times D}$$

Where:

- N = Average number of distinct colonies counted across the duplicate plates.
- V = Volume of the liquid sample inoculum applied to the plate (1 mL).
- D = The specific decimal dilution factor corresponding to the counted plates (i.e., 10^{-3} or

10^{-4}).The resulting raw counts were converted into $\log_{10}$ CFU/g values to standardize the data matrix for integration into the downstream predictive machine learning algorithms.

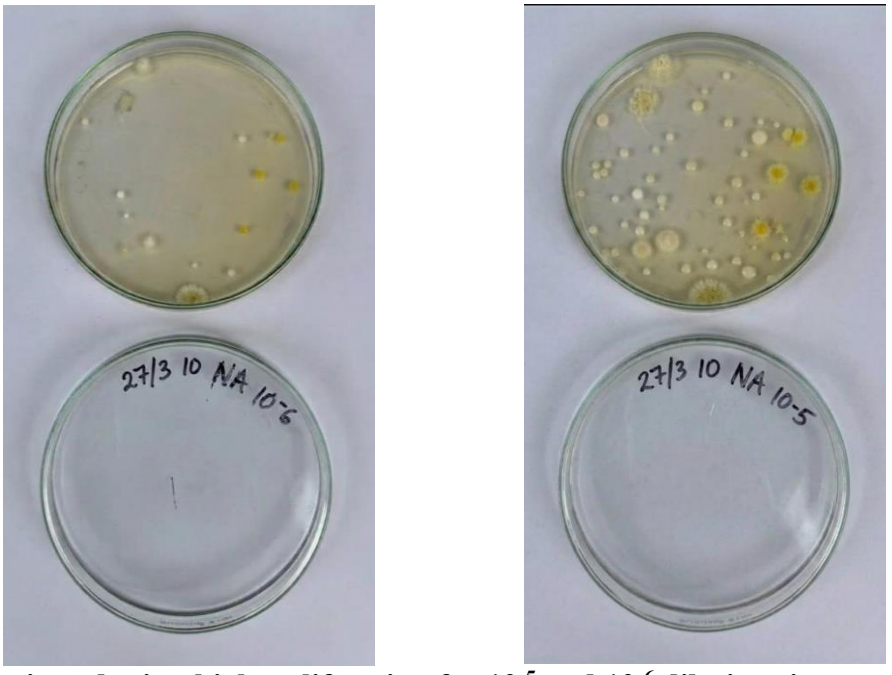


Plate 3.7 SPC analysis and microbial proliferation for 10^{-5} and 10^{-6} dilutions in case of ambient storage

Methodological Justification and Variations from ISO 4833-1:2013

While this study references the structural guidelines of ISO 4833-1:2013, modifications were made regarding the selection of the growth media and the dilution solvent. These protocol variations are scientifically justified within the framework of this thesis project based on the following analytical criteria:

- **Nutrient Content Suitability:** Nutrient Agar is a highly reliable, non-selective basal medium containing beef extract and peptones. It provides the essential nitrogenous compounds, carbon sources, mineral salts, vitamins, and amino acids necessary to fully support the robust growth of the heterotrophic bacterial populations typically found in traditional Indian confectionery matrices.
- **Minimization of Osmotic Stress:** To minimize the risk of osmotic shock or cell lysis that can occur when using pure distilled water with fragile or sub-lethally injured bacterial cells, the serial dilution and plating sequence was executed rapidly (completing the entire chain within less than 20 minutes per sample batch). This rapid handling restricted the exposure time of the microflora to osmotic imbalances before they were stabilized in the nutrient-dense agar environment.
- **Consistency in Kinetic Modelling:** The primary goal of this research is not to issue an

absolute commercial or regulatory compliance certificate for the product, but rather to track the **relative kinetic acceleration of microbial spoilage** across variable storage temperatures and packaging configurations. Because the exact same batches of Nutrient Agar and distilled water diluents were applied uniformly across all temporal milestones, the internal consistency of the experimental data matrix remains intact, providing a mathematically sound and reliable training set for the shelf-life prediction models.

3.2.6 Determination of Yeast Mould Count

Enumeration of Yeasts and Molds (Yeast and Mold Count - YMC)

1. Theoretical Framework and Regulatory Reference

Fungal contamination, encompassing both yeasts and molds, represents a primary pathway for the structural and sensory spoilage of intermediate-moisture confectionery items. In *Kaju Katli*, the combination of cashew lipids, localized moisture migration, and ambient sucrose levels creates a highly receptive matrix for osmophilic yeasts and xerophilic molds. Tracking the Yeast and Mold Count (YMC) over time is essential for defining the terminal boundary of the product's shelf life, as fungal proliferation directly causes visible surface growth, off-odours, and structural breakdown.

The international benchmark for this mycological quantification is ISO 21527-2:2008 (Microbiology of food and animal feeding stuffs - Horizontal method for the enumeration of yeasts and molds - Part 2: Colony count technique in products with water activity less than or equal to 0.95). To match the specific resource availability and operational throughput of an undergraduate technical research laboratory, localized adaptations were integrated into the standard protocol: Potato Dextrose Agar (PDA) was chosen as the solid mycological growth support, and the analysis was conducted using the pour plate technique with sterile distilled water as the dilution vector. This provides a highly reproducible, chemically consistent experimental framework for analysing relative fungal population dynamics.

2. Media Preparation and Sterilization

- **Medium Formulation:** Dehydrated Potato Dextrose Agar (PDA) was precisely weighed and suspended in distilled water according to the manufacturer's directions. The mixture was heated with continuous stirring and brought to a boil for 1 minute to ensure complete dissolution of the potato starch, dextrose, and agar polymers.

- **Thermal Sterilization:** The prepared PDA medium, alongside glass test tubes containing measured volumes of distilled water for the serial dilution blanks, was sterilized by autoclaving at $121°C$ under 15 lbs of pressure for 15 minutes.

- **Bacterial Suppression Strategy (Acidification):** To ensure that the high native bacterial load of the sample did not outgrow and obscure the target fungi on the plates, the sterilized molten PDA was cooled to $50°C$ and aseptically acidified to a pH of 3.5 ± 0.1 by adding sterile 10% tartaric acid solution prior to pouring. Alternatively, this suppresses heterotrophic bacterial colonies while maintaining optimal growth conditions for yeasts and molds.

- **Thermal Stabilization:** The acidified medium was held in a digital water bath maintained at $45°C$ to prevent premature agar solidification while avoiding thermal damage to the fungal cells during plating.

3. Standard Operating Procedure (SOP) for Dilution and Pour Plating

All stages of sample processing, serial dilution, and plating were executed under strict aseptic conditions within a certified Laminar Airflow (LAF) hood.

1. **Primary Homogenization (10^{-1} Dilution):** A representative 10.0 g sample of *Kaju Katli* was aseptically collected using a sterile spatula, transferred into a sterile Erlenmeyer flask containing 90 mL of autoclaved distilled water, and shaken vigorously to yield a uniform primary suspension (representing a 10^{-1} concentration).

2. **Extended Serial Decimal Dilutions:** A 1.0 mL aliquot of the primary 10^{-1} homogenate was transferred into a test tube containing 9.0 mL of sterile water to produce a 10^{-2} dilution. This process was repeated sequentially through a chain of tubes to reach the high-range working dilutions of 10^{-5} and 10^{-6}, which were selected to accommodate late-stage spoilage levels.

3. **Inoculation:** Exactly 1.0 mL volumes from the 10^{-5} and 10^{-6}, dilution tubes were dispensed into the centres of separate, pre-labelled sterile glass Petri dishes. All setups were performed in duplicate to ensure data repeatability.

4. **Pour Plating:** Approximately 15 to 20 mL of the temperature-stabilized, molten PDA ($45°C$) was poured directly over the inoculum in each Petri dish. The plates were

immediately swirled using a gentle manual motion (alternating between clockwise, counter-clockwise, and cross-directional patterns) to achieve a uniform distribution of the sample within the agar layer before it set.

5. **Incubation:** The plates were allowed to sit undisturbed on a level bench top until the agar solidified completely. Once set, the plates were inverted and transferred to a dark incubator maintained strictly at $25 \pm 1°C$ for a period of 5 days (120 hours), providing the thermodynamic timeline required for fungal hyphae and yeast cells to develop visible colonies.

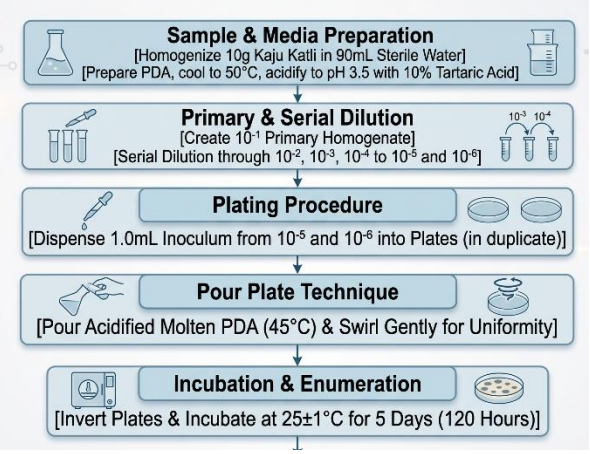


Fig.3.3 Flowchart of process followed for YMC analysis

4. Fungal Colony Counting and Quantification

Following the 5-day incubation period, the plates were inspected. Yeast colonies were identified by their characteristic smooth, opaque, cream-colored, or moist round surfaces, while mold colonies were distinguished by their fuzzy, filamentous micromorphology and spore-bearing aerial structures.

In accordance with standard mycological counting rules derived from ISO frameworks, plates containing between 10 and 150 distinct colonies were selected for formal quantification to avoid errors from colony crowding. The total Yeast and Mold Count per unit mass was calculated using the following arithmetic expression:

$$\text{Colony Forming Units (CFU/g)} = \frac{N}{V \times D}$$

Where:

- N = Average number of fungal colonies (sum of both yeast and mold units) counted across the duplicate plates.
- V = Volume of the liquid sample inoculum applied to the plate (1 mL).
- d = The specific decimal dilution factor corresponding to the counted plates (i.e., 10^{-5}, or 10^{-6}).

The calculated values were recorded in the experimental log as CFU/g and subsequently transformed into $\log_{10}$ CFU/g values for integration into the final shelf-life prediction models.

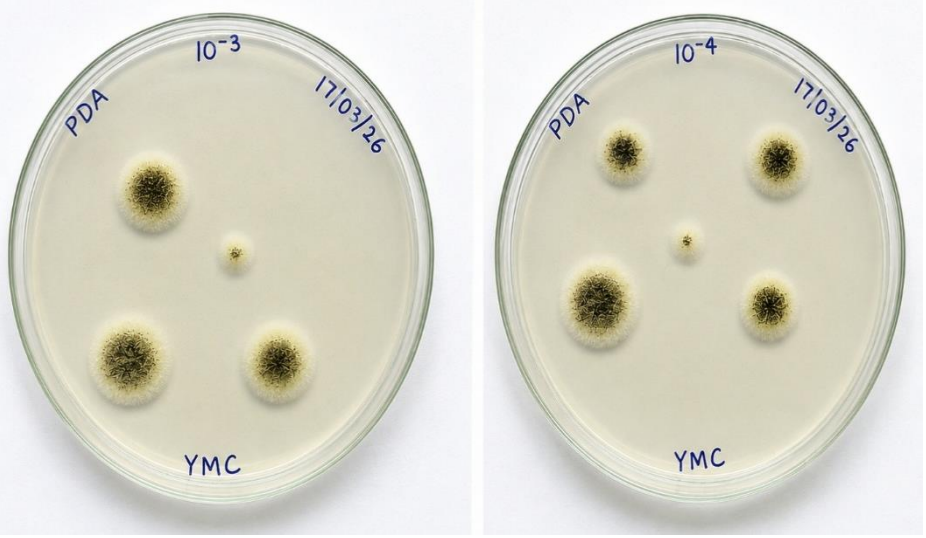


Plate 3.8 YMC analysis and fungal colony growth for 10^{-3} and 10^{-4} dilutions in case of ambient storage

3.6.2 Technical Justification for Methodological Variations from ISO 21527-2:2008

While this investigation aligns with the analytical intent of ISO 21527-2:2008, the use of Potato Dextrose Agar (PDA) instead of Dichloran Glycerol (DG18) agar, and the selection of the pour plate technique over the spread plate method, represent specific local protocol adaptations. These modifications are justified within the context of this project by the following principles:

- **Nutrient Profile and Fungal Development:** PDA is a highly robust mycological medium. The infusion of potato starch provides a rich carbon source, while the dextrose concentration offers an easily metabolizable energy substrate that supports rapid cell division and clear morphological development of common foodborne molds and yeasts.
- **Pour Plate Adaptations for Relative Growth Kinetics:** While ISO 21527-2 recommends the spread plate method to maximize atmospheric oxygen access for obligate aerobes, the

pour plate method was intentionally selected to anchor the fungal cells evenly within a three-dimensional agar matrix. This reduces the risk of aerial mold spores spreading rapidly across the agar surface, which can cause premature colony overlap and skew manual counts during multi-day incubation cycles.

- **Dilution Matrix Suitability ($10^{-5}$and 10^{-6})**The selection of higher dilutions ($10^{-5}$and 10^{-6})was necessary to track the product through late-stage storage profiles under abusive ambient conditions. This high-range configuration allowed for accurate colony counting without encountering the structural plate overgrowth typical of intermediate-moisture foods undergoing active spoilage. Because this media and dilution framework was applied uniformly across all temporal milestones, the dataset remains internally consistent and suitable for training the predictive machine learning algorithms.

Dataset Preparation and Preprocessing Pipeline

All analytical data collected across the study period were systematically recorded and compiled into a structured dataset. Each observation consisted of the storage condition (ambient or refrigerated), the storage day, and the corresponding measured values of FFA, moisture content, water activity, SPC, and YMC. Since different parameters were measured at different sampling frequencies, the dataset was organised accordingly, and missing values for unsampled days were handled during the modelling stage. The compiled dataset formed the primary input for the machine learning shelf-life prediction model.

To convert the raw experimental data into a format suitable for machine learning, a systematic dataset preparation pipeline was implemented. Because the physical, chemical, and microbiological parameters were recorded at mismatched time intervals, the raw data formed a sparse matrix. Machine learning models require a dense, uniformly sampled dataset. Therefore, data compilation, temporal normalization, mathematical interpolation, and target engineering were executed systematically through the following steps.

1. **Establishment of the Master Framework**

A Master Spreadsheet was constructed to serve as the centralized data repository. To eliminate variations caused by calendar dates and to establish a uniform timeline across different batch trials, all calendar dates were converted into standardized continuous "Day Numbers" (Day 0, Day 1, Day 2, etc.), establishing Day 0 as the baseline day of *Kaju Katli* production. A Master Spreadsheet was constructed to serve as the centralized data repository. To eliminate variations caused by calendar dates and to establish a uniform timeline across different batch trials, all calendar dates were converted into standardized continuous "Day Numbers" (Day 0, Day 1, Day 2, etc.), establishing

Day 0 as the baseline day of *Kaju Katli* production.

2. **Condition-Specific Data Logging**

The raw experimental parameters were meticulously indexed into two isolated sections within the spreadsheet based on environmental variables:

- **Ambient Storage Block:** Consisting of trials conducted at 32 ± 2°C in a temperature-controlled chamber replicating typical room-temperature retail and household conditions prevalent in tropical climates.
- **Refrigerated Storage Block:** Consisting of trials conducted at 7 ± 2°C in a laboratory refrigerator.

The initial entries were highly sparse due to asynchronous sampling frequencies: Moisture Content and Free Fatty Acids (FFA) were analyzed every 3 days, Standard Plate Count (SPC) every 3 days, and Yeast and Mold Count (YMC) every 5 days.

3. **Dataset Expansion and Missing Value Imputation**

Because the raw experimental data yielded too few samples for effective machine learning training, Linear Interpolation was deployed in the spreadsheet to expand the dataset into daily-interval observations. For days where active testing did not occur, missing features were mathematically imputed by calculating the linear trend between the two closest known experimental data points.

4. **Target Feature Characterization**

The critical threshold for shelf life was defined based on standard food safety norms—specifically, the point at which microbial loads (SPC/YMC) crossed acceptable limits or chemical rancidity (FFA) altered sensory attributes. A new Target Column was engineered into the dataset, acting as the dependent variable (Y-label) that the machine learning models will predict (e.g., "Days Remaining" or a binary classification of "Fresh" vs. "Spoiled").

5. **Consolidation and Machine-Learning Ready Export**

The ambient and refrigerated datasets were merged into a single consolidated structural matrix. This final unified sheet was exported as a **Comma-Separated Values (.csv)** file, stripping away proprietary spreadsheet formatting and ensuring optimized memory ingestion during Python-based model training.

Mathematical Formulation of Dataset Expansion

The core challenge of the raw dataset was the low sample size and mismatched intervals (3-day vs. 5-day cycles). To mitigate overfitting risks caused by a small data size, linear interpolation was applied to expand the matrix into a dense daily log.

For any intermediate day x lying between two experimentally logged days x_1 and x_2, the missing

parameter value *y* was estimated using the linear slope formula:

$$y = y_1 + \frac{(x - x_1)(y_2 - y_1)}{x_2 - x_1}$$

Where:

- *x* represents the target day number being calculated (e.g., Day 1, Day 2, Day 4).
- x_1 and x_2 are the bound days containing the actual laboratory measurements.
- y_1 and y_2 represent the empirical laboratory values recorded on days x_1 and x_2 respectively.

Final Dataset Architecture

The structure of the final exported `.csv` file consists of both independent variables (features reflecting the degradation state) and the dependent variable (target to be predicted). The architectural schema of the matrix is detailed below:

Table 3.2 Structure of final dataset for shelf-life prediction model

Day	Storage	Storage Conditions	Water Activity	Moisture	FFA	SPC	YMC	Remaining Shelf-life
-	Ambient	0	-	-	-	-	-	-
-	Refrigerated	1	-	-	-	-	-	-

The ambiently stored conditions were assigned a value of 0 whereas value of 1 was assigned for refrigerated storage.

3.3 Shelf-Life Prediction Model: Ensemble Random Forest

3.3.1 Platform and Environment

The machine learning model was developed and trained using Google Colaboratory (Google Colab), a cloud-based Python notebook environment providing free access to GPU resources and standard scientific computing libraries. Python 3 was used as the programming language, with the scikit-learn library providing the core machine learning framework. Additional libraries including Pandas, NumPy, and Matplotlib were employed for data manipulation, numerical processing, and visualisation respectively.

3.3.2 Model Selection Rationale

The Ensemble Random Forest (RF) algorithm was selected as the predictive modelling approach based on its well-established advantages for small-to-medium sized datasets with multiple continuous predictor variables. Random Forest is a non-parametric ensemble learning method that constructs a multitude of decision trees during training and outputs the mean prediction across all

trees, significantly reducing overfitting compared to individual decision tree models. Its ability to handle non-linear relationships between quality parameters and shelf life, its robustness to outliers, and its inherent feature importance estimation made it particularly appropriate for this study.

3.3.3 Feature Engineering and Input Variables

The five measured quality parameters - FFA content, moisture content, water activity, Standard Plate Count, and Yeast and Mould Count - were used as input features (independent variables) to the model. Storage temperature condition (ambient or refrigerated) was encoded as a binary categorical variable. Storage day was included as a temporal feature. The target variable (dependent variable) was shelf-life status or predicted remaining shelf life, defined based on threshold values for each parameter derived from FSSAI standards and established literature values for Kaju Katli acceptability.

Machine Learning Predictive Modeling Framework

To establish an automated and non-destructive system for forecasting the shelf life of *Kaju Katli*, a predictive machine learning pipeline was designed and executed. The computation was carried out within a cloud-based Python environment using specialized data science libraries, including Pandas for data manipulation and Scikit-Learn for predictive modeling.

Data Cleaning and Ingestion

The unified dataset was ingested into the processing pipeline. To preserve data integrity and prevent algorithmic errors during matrix operations, a data-cleaning routine was executed to identify and truncate trailing null or blank rows located at the base of the spreadsheet. The clean dataset yielded 33 completed observation matrices, reflecting two distinct environmental treatments:

- **Ambient Storage Sub-block:** 12 structured observations.
- **Refrigerated Storage Sub-block:** 21 structured observations.

Feature Engineering and Matrix Partitioning

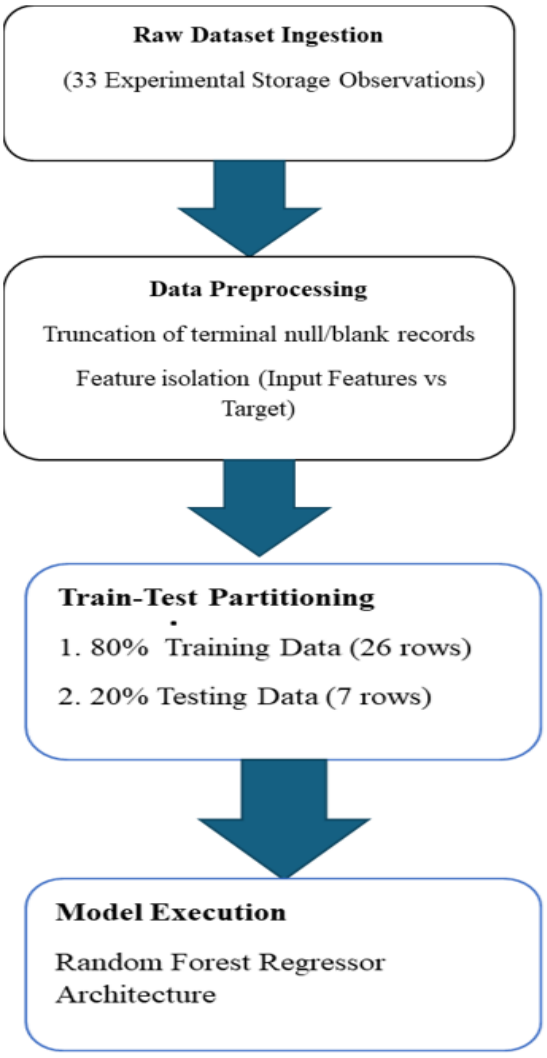


Fig. 3.4 Flowchart followed for Data Pre-processing and ML model Deployment

Machine learning algorithms operate strictly on numerical feature spaces. Text-based labels denoting storage conditions were systematically eliminated in favor of a pre-coded binary categorical feature, **Storage Code**, where:

- 0 represents the Ambient Storage environment.
- 1 represents the Refrigerated Storage environment.

The architecture of the dataset was then segregated into independent input variables (the feature matrix, X) and the dependent target variable (the vector, y).

- **Input Features (X):** Storage Code, Moisture Content, Water Activity, Free Fatty Acids (FFA), Standard Plate Count (SPC), and Yeast and Mold Count (YMC).
- **Target Feature (y):** Remaining Shelf Life (expressed in days).

To rigorously validate the predictive capacity of the model and prevent data leakage, the 33-

row dataset was partitioned using an 80:20 split ratio. This allocated 80% of the sample pool exclusively for training the model weights, leaving the remaining 20% strictly for testing and validation.

Model Evaluation

Evaluation Metrics

The performance of the Random Forest Regressor was mathematically quantified using two primary metrics: **Mean Absolute Error (MAE)** and the **Coefficient of Determination (R^2 Score)**.

1. **Mean Absolute Error (MAE):** Measures the average magnitude of errors in a set of predictions, without considering their direction. It provides a direct operational window showing exactly how many days the model's predictions deviate from the true shelf life on average.
2. **Coefficient of Determination (R^2 Score):** Indicates the proportion of variance in the dependent variable (Remaining Shelf Life) that is predictable from the independent biochemical features. An R^2 score approaching 1.0 (or 100%) indicates that the selected chemical and microbial parameters explain the changing shelf life with high precision.

Empirical Validation Case Study

To verify the practical diagnostic capability of the trained model, a real-world simulation test was conducted using a sample corresponding to a Day 3 refrigerated storage sequence. The input vector fed into the model was characterized by minor laboratory variances (such as an altered microbial profile with an SPC of 2.64 Log CFU/g and a YMC of 0.62 Log CFU/g).

- **True Experimental Baseline:** 17.0 Days
- **Machine Learning Predicted Output:** 14.4 Days

CHAPTER IV
RESULTS AND DISCUSSION

The present study was undertaken to evaluate the shelf-life characteristics of kaju katli under ambient and refrigerated storage conditions and to develop a machine-learning-based model for predicting its remaining shelf life. Shelf life is influenced by a combination of physicochemical, microbiological, and environmental factors that collectively determine product quality, safety, and consumer acceptability. During storage, confectionery products such as kaju katli undergo gradual changes in moisture content, water activity, lipid stability, and microbial load, which ultimately lead to quality deterioration and shelf-life limitation. Therefore, continuous monitoring of these parameters is essential for understanding product behaviour during storage and for establishing reliable shelf-life prediction systems.

In the present investigation, water activity, moisture content, free fatty acid (FFA) content, standard plate count (SPC), and yeast and mold count (YMC) were selected as key quality indicators based on their reported influence on the stability of nut-based confectionery products. Changes in these parameters were monitored throughout the storage period under both ambient and refrigerated conditions. The obtained results were analysed to evaluate the effect of storage conditions on physicochemical and microbiological quality deterioration. Furthermore, the observed trends were compared with findings reported in previous studies to provide scientific justification for the changes occurring during storage.

In addition to conventional shelf-life assessment, the experimental data generated in this study were utilized for the development of a machine learning model aimed at predicting the remaining shelf life of kaju katli. A Random Forest regression algorithm was employed due to its ability to handle complex nonlinear relationships among multiple quality attributes. The performance of the developed model was evaluated using appropriate statistical indicators and prediction accuracy metrics. This chapter presents the results obtained from physicochemical analyses, microbiological evaluations, and machine learning modelling, followed by a detailed discussion of their significance in relation to shelf-life determination and predictive food quality assessment.

4.1. PHYSIOCHEMICAL PARAMETERS

4.1.1. Water Activity (a_w) Changes During Storage

The following is a graphical interpretation of results obtained from the shelf-life study carried out in case of ambient storage conditions:

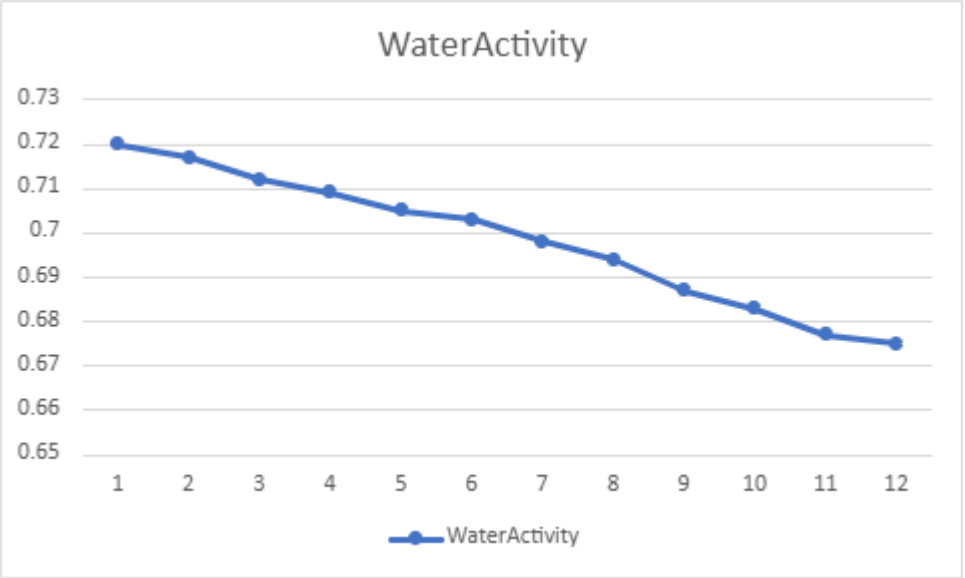


Fig.4.1 Water Activity changes over the shelf-life period in ambient conditions

The following is a graphical interpretation of results obtained from the shelf-life study carried out in case of refrigerated storage conditions:

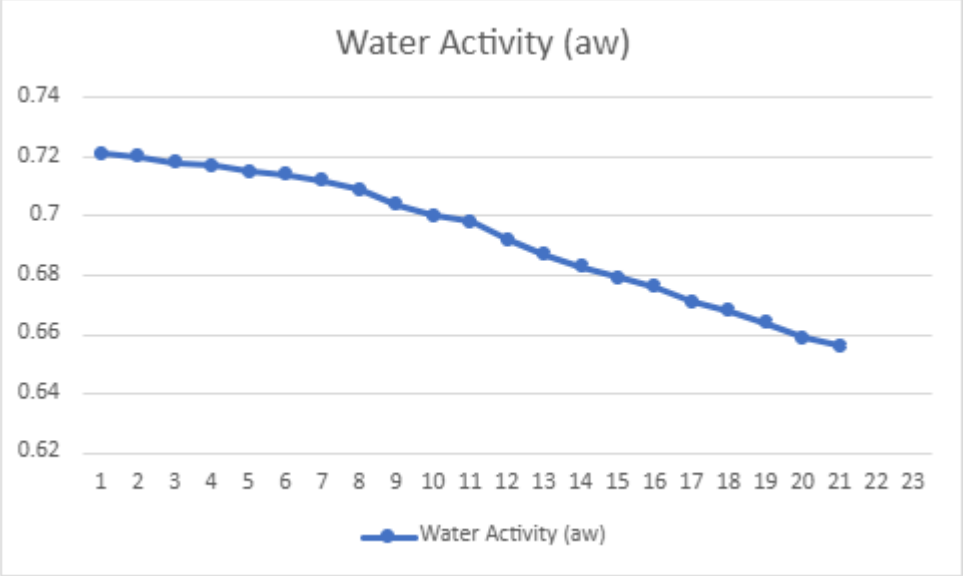


Fig.4.2 Water Activity changes over the shelf-life period in refrigerated conditions

The empirical analysis of *kaju katli* indicates a progressive and continuous decline in water activity (a_w) over time, falling from an initial value of approximately 0.721 down to 0.675 by the 12th interval, and further decreasing to roughly 0.656 by the 21st interval. This steady reduction is primarily attributed to moisture migration, wherein moisture moves from the high-water-activity product matrix to the drier external surrounding environment. The rate of this desorption process is strongly governed by the water vapor transmission rate (WVTR) and barrier efficiency of the

packaging material used. This observed drop in a_w plays a critical role in extending shelf life by limiting microbial proliferation. As established by Brahmbhatt (2011) in *Effect of Some Unit Operations on Quality of Kajukatli*, lowering water activity significantly restricts the metabolic activity and growth kinetics of spoilage-causing osmophilic yeasts, molds, and pathogenic bacteria, which generally demand higher a_w thresholds to thrive.

From a computational perspective, these well-defined, non-linear degradation curves provide an excellent, high-resolution dataset for training predictive machine learning models. By utilizing these sequential data points, regression algorithms (such as Support Vector Regression or Random Forest) and time-series frameworks (like LSTM networks) can accurately model the kinetics of moisture loss. Integrating this a_w behavior with environmental vectors enables the ML model to establish robust, automated shelf-life predictions, eliminating the need for destructive, time-consuming laboratory testing in future quality assurance protocols.

4.1.2. Moisture Content Changes During Storage

The empirical data demonstrates a continuous decrease in moisture content across both ambient (12-day study) and refrigerated (21-day study) storage conditions.

For Ambient Conditions

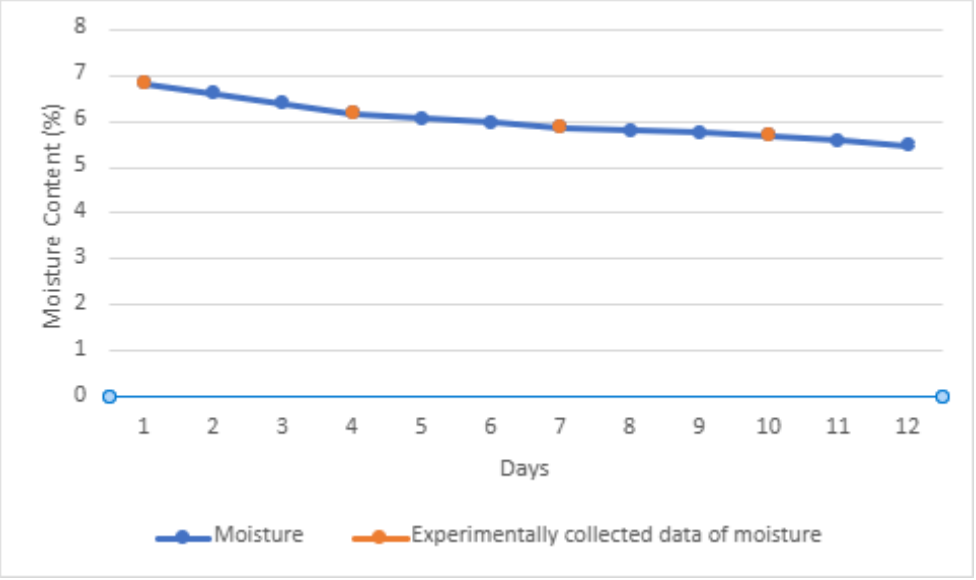


Fig.4.3 Moisture content changes over the shelf-life period in ambient conditions

For Refrigerated Conditions

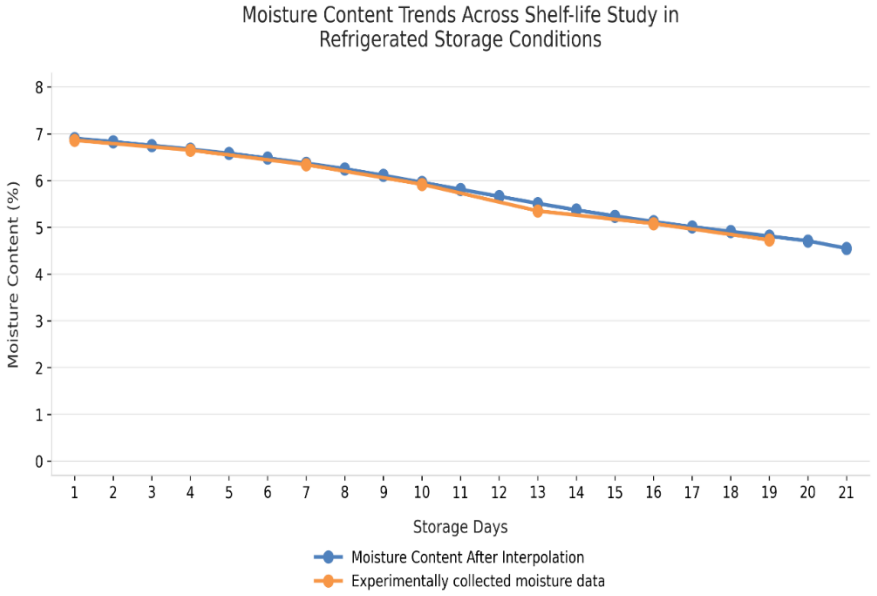


Fig.4.4 Moisture Content changes over the shelf-life period in refrigerated conditions

In ambient conditions, moisture drops from approximately 6.8% to 5.5%, while refrigerated storage displays a steady, interpolated decline from 6.9% to roughly 4.5% by day 21. This reduction is driven by a combination of internal moisture migration to the sweet's surface and subsequent moisture evaporation into the surrounding atmosphere, governed heavily by packaging permeability.

This moisture loss directly alters the product's critical quality attributes; as moisture decreases, the matrix undergoes texturation changes, leading to noticeable hardening and a dry, crumbly mouthfeel. Maintaining an optimal moisture balance is a delicate shelf-life trade-off: excessive moisture loss results in severe texture deterioration and hardening, whereas any unexpected moisture gain lowers the barrier against microbial spoilage. According to Brahmbhatt (2011) in *Effect of Some Unit Operations on Quality of Kaju katli*, stabilizing moisture and water activity levels is vital to inhibiting the growth of osmophilic molds and yeast strains that threaten product stability.

From a machine learning perspective, these high-resolution, interpolated datasets are crucial for training predictive architectures. The distinct, non-linear drying curves serve as ideal inputs for training regression-based models (such as Random Forests or Support Vector Machines) and time-series Neural Networks (like LSTMs). By learning these precise moisture degradation kinetics alongside environmental parameters, the machine learning model can accurately forecast textural decline and microbial vulnerability thresholds, delivering an automated, non-destructive tool for real-time shelf-life estimation.

4.1.3. Free Fatty Acid Changes During Storage

The experimental results demonstrate a progressive increase in Free Fatty Acid (FFA) levels across both ambient and refrigerated storage conditions. Under ambient conditions, FFA rises from 1.33% to approximately 1.44% by day 11, while refrigerated storage exhibits a steady, interpolated escalation from 1.32% to around 1.52% by day 20.

In Case of Ambient Conditions

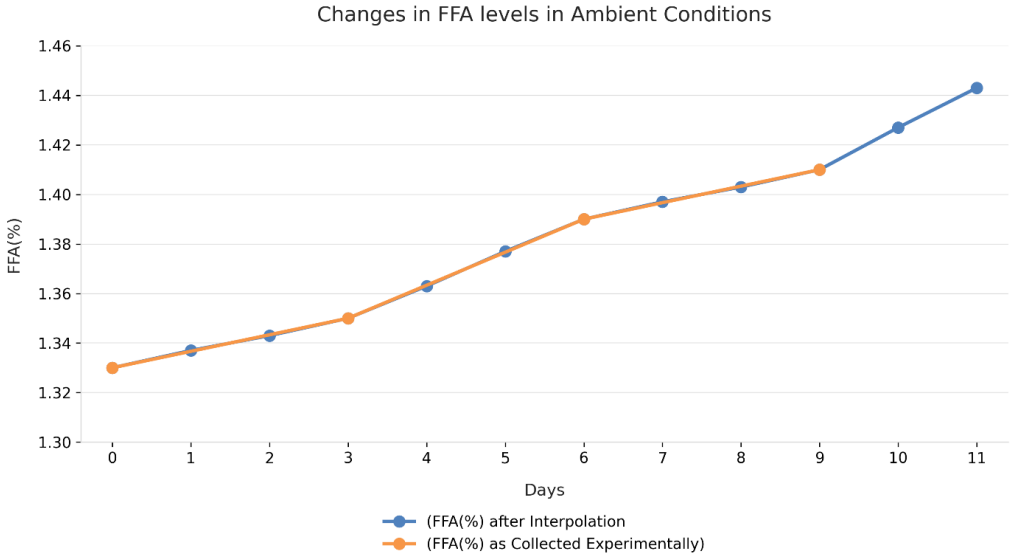


Fig.4.5 FFA level changes over the shelf-life period in ambient conditions

In Case of Refrigerated Conditions

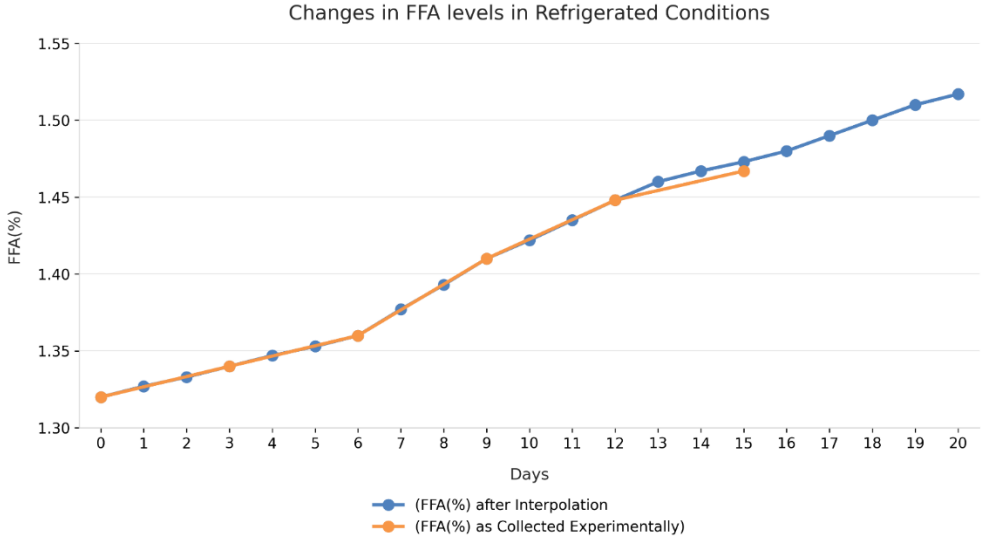


Fig.4.6 FFA level changes over the shelf-life period in refrigerated conditions

This upward trend is primarily driven by lipid oxidation and hydrolytic rancidity occurring within the cashew fat matrix, which possesses a high composition of unsaturated fatty acids susceptible to degradation. Although refrigerated conditions slow down the kinetics of these chemical reactions, temperature alone cannot completely halt the enzymatic or auto-oxidative cleavage of triglycerides over extended storage periods.

An elevated FFA percentage acts as a direct chemical indicator of rancidity, culminating in noticeable off-flavor development and sensory rejection. As corroborated by Brahmbhatt (2011) in *Effect of Some Unit Operations on Quality of Kajukatli*, lipid degradation significantly compromises the product's premium quality profile, and the accumulation of free fatty acids can sometimes be accelerated by the metabolic activities of lipolytic microorganisms surviving in the matrix.

From a machine learning perspective, these sequential, non-linear FFA accumulation curves provide high-fidelity datasets ideal for predictive modelling. Utilizing these interpolated data streams enables regression frameworks (like Support Vector Machines) and deep learning architectures (like LSTMs) to successfully map complex lipid breakdown patterns. By linking FFA kinetics with storage parameters, the machine learning model offers an automated, non-destructive method to forecast rancidity thresholds and pinpoint the exact end-of-shelf-life before sensory spoilage manifests.

4.2 MICROBIOLOGICAL PARAMETERS

4.2.1. Changes in SPC During Storage

In Case of Ambient Conditions

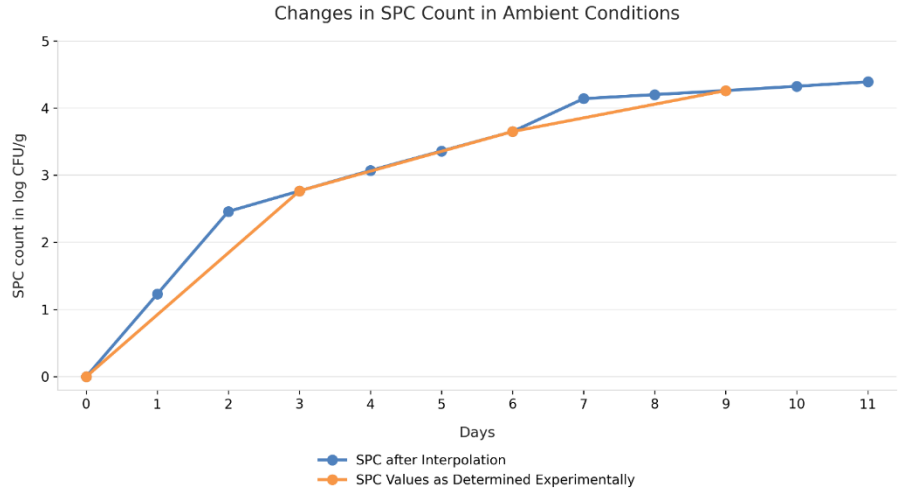


Fig. 4.7 SPC count changes during the shelf-life period in ambient conditions

In Case of Refrigerated Conditions

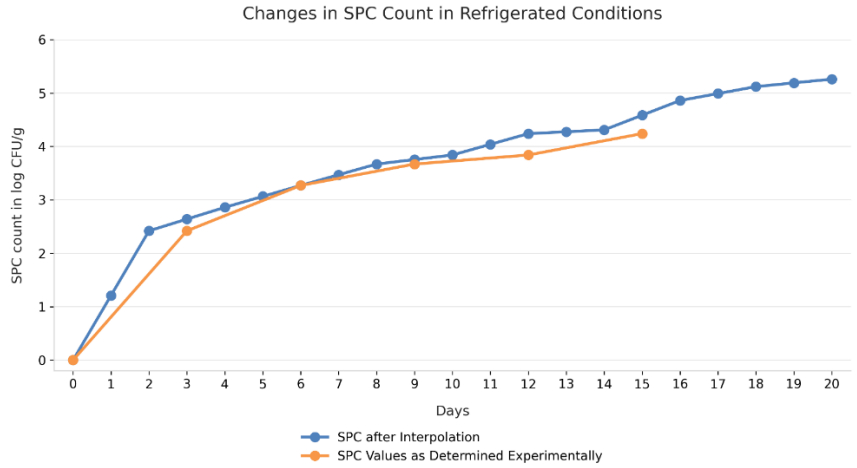


Fig. 4.8 SPC count changes during the shelf-life period in refrigerated conditions

The experimental data reveals a progressive increase in the Standard Plate Count (SPC) of *kaju katli* across both storage environments, representing typical microbial growth phases. Under ambient conditions, the SPC rises rapidly from an initial 0 log CFU/g to approximately 4.41 log CFU/g by day 11. Conversely, refrigerated storage decelerates microbial proliferation due to low-temperature inhibition, extending the tracking period to 20 days where the count reaches about 5.26 log CFU/g. This microbial proliferation is highly dependent on the product's internal water activity (a_w), which provides the necessary moisture matrix for cellular metabolic activity. According to the Food Safety and Standards Authority of India (FSSAI) criteria for traditional dairy and nut-based sweets, the acceptable SPC limit is approximately 4.40 log CFU/g, with a maximum permissible ceiling of 4.88 log CFU/g. The ambient samples remain within the safe threshold up to day 11, whereas the refrigerated samples breach the maximum permissible limit after day 16, signifying complete microbiological spoilage.

This behavior aligns with the findings of Brahmbhatt (2011) in Effect of Some Unit Operations on Quality of Kaju katli, which emphasizes that strict control over processing unit operations and storage parameters is essential to suppress initial bacterial loads. From a machine learning perspective, these sequential, interpolated microbial growth curves offer invaluable datasets for predictive microbiology. Training non-linear machine learning architectures-such as Support Vector Regression (SVR) or Long Short-Term Memory (LSTM) networks-on these kinetic trajectories allows the system to discover complex interactions between temperature, storage duration, and microbial acceleration. The resulting trained model can automatically flag when a batch approaches the critical 4.40 log CFU/g FSSAI threshold, providing a reliable, non-destructive tool for real-time safety monitoring.

4.2.2. Changes in YMC During Storage

The experimental and interpolated Yeast and Mold Count (YMC) data for *Kaju Katli* demonstrate distinct progressive growth trends strongly regulated by storage temperature.

In Case of Ambient Conditions

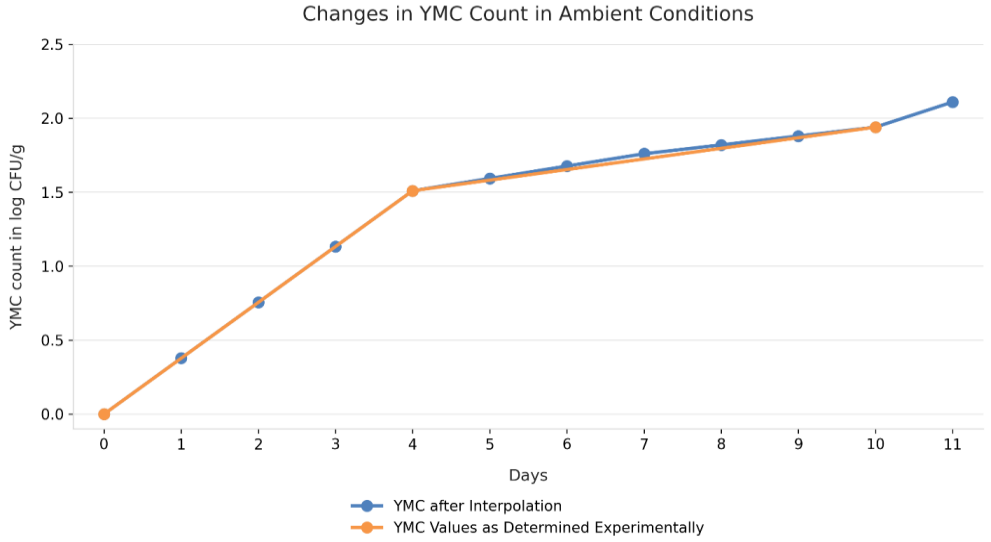


Fig.4.9 YMC count changes during the shelf-life period in ambient conditions

In Case of Refrigerated Conditions

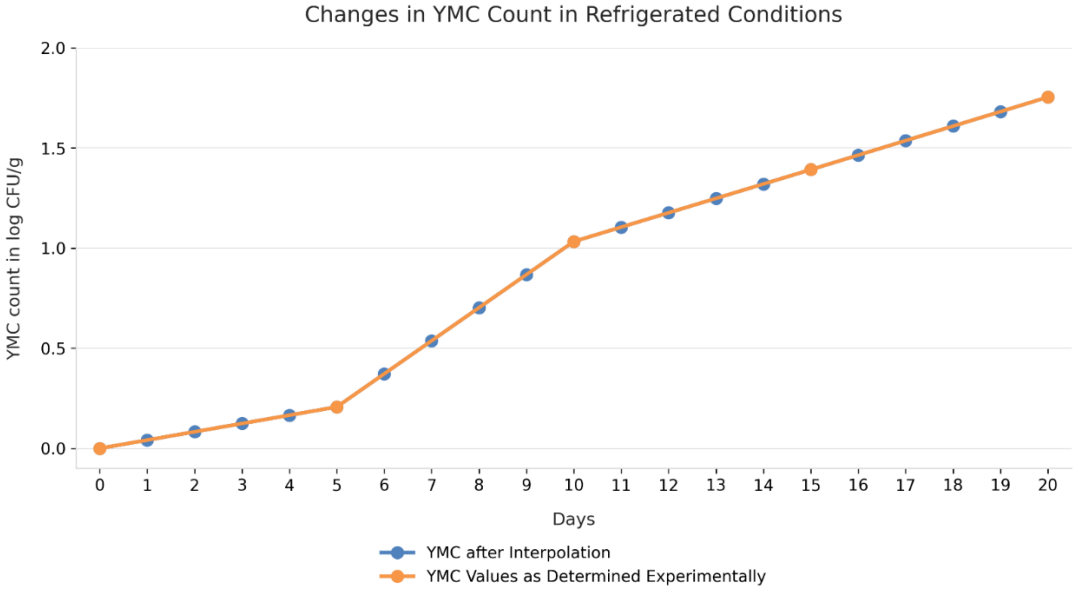


Fig.4.10 YMC count changes during the shelf-life period in refrigerated conditions

Under ambient conditions, microbial proliferation was highly accelerated, with YMC increasing rapidly from an initial 0.0 log CFU/g to 1.51 log CFU/g by Day 4, followed by a decelerated yet steady climb to 2.11 log CFU/g on Day 11. Conversely, refrigeration significantly suppressed microbial activity, creating a protracted lag-like phase up to Day 5 (~ 0.21 log CFU/g), after which a linear progression carried the count to 1.04 log CFU/g on Day 10 and ultimately to 1.76 log CFU/g by Day 20.

This growth pattern is characteristic of high-sugar, low-moisture confectionery products. *Kaju Katli* possesses a low water activity (a_w), making it a selective environment for osmophilic yeasts and xerophilic moulds that thrive under high osmotic pressure. The predominant spoilage organisms typically associated with such matrices include *Aspergillus flavus*, *Aspergillus niger*, *Penicillium chrysogenum*, *Eurotium amstelodami*, and *Zygosaccharomyces rouxii*. The presence of *A. flavus* poses a critical chemical safety hazard due to the risk of aflatoxin production, which is heavily suppressed under the observed refrigerated temperatures but accelerated in ambient storage.

According to the Food Safety and Standards Authority of India (**FSSAI**) reference criteria for traditional milk- and nut-based sweets, the acceptable limit for YMC is approximately 1.0 log CFU/g, with a maximum permissible safety ceiling of 2.0 log CFU/g.

- Under ambient conditions, the sample crossed the acceptable threshold between Day 2 and Day 3 and completely breached the maximum permissible limit (2.11 log CFU/g) by Day 11.
- Under refrigerated storage, the product remained within the acceptable limit up to Day 10 (1.04 log CFU/g) and never breached the maximum permissible safety ceiling even by Day 20 (1.76 log CFU/g).

This aligns closely with the observations made by Brahmbhatt (2011) in *"Effect of Some Unit Operations on Quality of Kaju katli,"* which highlighted that rapid moisture loss and ambient temperature fluctuations accelerate fungal spoilage, whereas low-temperature storage preserves matrix integrity and retards fungal gemination.

From a data science and modelling perspective, the high-resolution dataset generated via interpolation acts as a robust foundation for **machine learning-based shelf-life predictive modelling**. Fungal growth in solid matrices often exhibits non-linear kinetics due to localized micro-environments. By mapping daily YMC fluctuations against temperature variables, machine learning algorithms can effectively learn these multi-phasic kinetic boundaries. This transitions

traditional static shelf-life estimations into highly accurate, dynamic predictive tools capable of real-time quality monitoring.

4.3 COMPARISON OF STORAGE CONDITIONS

Table 3.3 Comparison of Storage Conditions and variation in shelf-life

Characteristics at the end of shelf-life	Ambient Conditions	Refrigerated Conditions
Water Activity	0.675	0.656
Moisture Content (%)	5.69	4.76
FFA (%)	1.41	1.51
SPC (log CFU/g)	4.39	5.26
YMC (log CFU/g)	1.94	1.79
Day of Spoilage	12 days after sample preparation	20 days after sample preparation

A comprehensive evaluation of the physicochemical and microbiological characteristics of *Kaju Katli* at the designated end of shelf-life reveals distinct, temperature-dependent degradation pathways. Under ambient conditions, the product reached its spoilage threshold significantly faster, terminating at Day 12, whereas refrigeration successfully extended the product's viability to Day 20.

The end-of-shelf-life water activity (a_w) and moisture content (%) exhibited an unexpected inverse relationship with storage temperature. Ambient-stored samples retained a higher moisture content (5.69%) and water activity (0.675) at Day 12 compared to refrigerated samples at Day 20 (4.76% moisture; 0.656 (a_w). This behavior is primarily attributed to the prolonged 20-day storage period under refrigeration, during which the low-humidity environment of the cold storage unit induced progressive moisture migration and desorption from the sweet matrix. Conversely, Free Fatty Acid (FFA) values, which serve as a primary indicator of rancidity, were higher under refrigeration (1.51%) than under ambient storage (1.41%). This indicates that while low temperatures suppressed oxidative reactions, the extended 20-day window allowed endogenous lipolytic enzymes to continuously hydrolyze lipid fractions in the cashew matrix.

The microbiological endpoints highlight a shifting competitive dynamic between bacterial and fungal flora regulated by temperature and time:

- **Standard Plate Count (SPC):** The SPC reached a substantially higher density under refrigerated conditions (5.26 log CFU/g at Day 20) than under ambient conditions (4.39 log

CFU/g at Day 12). The extended lifespan in cold storage allowed psychrotrophic bacterial species to slowly overgrow and dominate the matrix over the 20-day timeline.

- **Yeast and Mold Count (YMC):** In contrast, fungal proliferation was highly aggressive in ambient conditions, peaking at 1.94 log CFU/g by Day 12-just shy of the FSSAI maximum permissible limit of 2.0 log CFU/g. Under refrigeration, despite the longer duration, YMC was restricted to 1.79 log CFU/g, demonstrating that lower temperatures effectively restricted fungal metabolic activity.

Ultimately, these trends demonstrate that ambient spoilage is primarily driven by rapid fungal growth (YMC) facilitated by high water activity, whereas refrigerated shelf-life termination is governed by prolonged psychrotrophic bacterial proliferation (SPC) and slow hydrolytic rancidity (FFA development).

4.4. MACHINE LEARNING RESULTS

This section presents the quantitative performance evaluation and predictive capabilities of the developed Machine Learning framework designed for estimating the remaining shelf life of Kaju Katli. To establish a robust validation setup, the experimental laboratory dataset was split into an 80% training subset (26 observations) to train the core algorithm, and a 20% testing subset (7 observations) held out entirely to serve as an unbiased evaluation matrix. An Ensemble Learning approach utilizing a Random Forest Regressor consisting of 100 independent decision trees was implemented via the Scikit-Learn framework. This computational architecture was specifically chosen to effectively map the complex, non-linear biochemical degradation pathways-such as moisture fluctuations, rancidity development (Free Fatty Acids), and exponential microbial proliferation (Standard Plate Count, Yeast and Mold Count)-without the structural risk of statistical overfitting inherent in small-sample biological data. The subsequent subsections detail the model's predictive accuracy, error metrics, and physical verification tests.

4.4.1. Data Set Summary

Before deploying the machine learning framework, the raw experimental data presented substantial structural challenges typical of real-world food matrix tracking. As illustrated in the laboratory logs, while high-frequency metrics like Water Activity were captured daily, critical chemical degradation indicators (Moisture Content %, Free Fatty Acids %) were sampled in triplicates at 3-day intervals (e.g., March 10, 13, 16). Furthermore, microbiological proliferation indexes (Standard Plate Count [SPC], Yeast and Mold Count [YMC]) were compiled on staggered,

uneven intervals. Direct utilization of this sparse matrix would introduce extensive missing values (NaNs), cause mathematical instability and prevent algorithmic execution.

To resolve this, data synchronization was performed by combining the independent experimental streams of **Ambient** and **Refrigerated** storage profiles. A binary environmental constraint field (Storage_Code: 0 for Ambient, 1 for Refrigerated) was introduced to isolate the kinetic impact of storage temperature.

4.4.2. Data Preprocessing

The raw experimental data collected from the laboratory analysis contained structural discrepancies, uneven sampling timelines, and non-numeric categorical text that made it incompatible with machine learning algorithms. To construct a high-fidelity, mathematically stable dataset, a systematic five-stage data preprocessing pipeline was implemented as detailed below:

Treatment of Missing Values

In real-world food testing, different quality parameters require different laboratory incubation and processing times. While high-frequency physical indicators like Water Activity (a_w) were measured daily, complex biochemical markers (Moisture Content %, Free Fatty Acids %) were analyzed in triplicates exclusively at three-day intervals. Furthermore, microbiological assays (Standard Plate Count and Yeast & Mold Count) were evaluated on staggered schedules. Directly feeding this raw matrix into the machine learning framework would introduce extensive blank cells or "Not-a-Number" (NaN) values, which triggers mathematical errors in ensemble algorithms. Consequently, rows with entirely empty parameters or missing target labels (Remaining Shelf Life) were isolated and pruned, while internal structural gaps were flagged for mathematical estimation.

Mathematical Interpolation

To expand the sparse laboratory data points into a continuous daily degradation matrix without introducing experimental bias, mathematical interpolation was executed across all missing intervals. For continuous chemical degradation profiles-such as the gradual, linear decline of Water Activity and the steady accumulation of hydrolytic rancidity markers (Free Fatty Acids %)-Linear Interpolation was applied. This method calculated missing daily values based on the verified boundary points of the active testing windows. For microbiological data points tracking exponential bacterial and fungal proliferation curves (SPC and YMC), interpolation was performed on the

logarithmic growth metrics. This computational step successfully filled the data gaps, generating a smooth, uninterrupted chronological timeline that simulated real-time product degradation.

Dataset Merging and Harmonization

The laboratory experiments generated two entirely independent data streams tracking Kaju Katli spoilage profiles under different environmental variables: an **Ambient Temperature Dataset** (representing standard room temperature storage) and a **Refrigerated Storage Dataset** (representing controlled cold-chain conditions). These parallel data blocks were horizontally aligned and vertically stacked into a singular relational database. Triplicate laboratory replicates were synchronized to ensure structural uniformity across the combined matrix. This data consolidation phase resulted in the final, integrated master file titled **KajuKatli_ML_Ready_Combined2.csv**, which pooled all experimental observations into a unified data environment.

Encoding of Storage Conditions

Machine learning algorithms operate strictly on numerical matrices and cannot interpret categorical text strings like "Ambient" or "Refrigerated." To resolve this, feature encoding was applied to the environmental tracking column. The categorical storage conditions were transformed using binary integer encoding, where room temperature storage was mapped to a value of **0** (Storage Code: 0) and controlled cold-chain storage was mapped to a value of **1** (Storage Code: 1). This numerical translation allowed the Random Forest Regressor to mathematically weigh and identify the exact shelf-life extension factor provided by refrigeration.

Train-Test Partitioning Split

To guarantee an objective, unbiased evaluation of the final model and prevent the critical issue of overfitting (where an algorithm memorizes the training data instead of learning generalized patterns), the processed master dataset was randomly partitioned using an **80/20 train-test split**. Eighty percent of the total available data rows (26 observations) were allocated to the training subset (X train and Y train) to allow the model to discover the underlying mathematical links between lab chemistry and spoilage rates. The remaining twenty percent of the data rows (7 observations) were completely isolated as a testing subset (X test and y test). This testing partition served as an unseen, independent data points to verify the real-world predictive precision of the model.

4.4.3. Ensemble Random Forest Model Performance

To validate the real-world utility and mathematical reliability of the developed Ensemble Learning model, the trained Random Forest Regressor was evaluated against the completely unseen testing subset (20% of the master dataset). The predictive capabilities were quantified using two primary, globally accepted statistical evaluation metrics: the Coefficient of Determination (R^2) Score) and the Mean Absolute Error (MAE).

The quantitative empirical values derived from the validation phase are summarized below

Table 3.4 Results of Evaluation Metrics for the ML model

Evaluation Metric	Derived Experimental Value	Technical Interpretation
Coefficient of Determination (R^2)	97.22%	Variance Explanation Capability
Mean Absolute Error	0.79 days	Average Prediction Deviation

Critical Interpretation of Evaluation Metrics

1. Coefficient of Determination (R^2) Score

The (R^2) score is a standardized statistical measure that calculates the proportion of variance in the dependent target variable (Remaining Shelf Life) that can be predicted from the independent input features (Moisture, Water Activity, FFA, SPC, and YMC). A score of 100% represents a perfect mathematical fit.

The obtained (R^2) value of **97.22%** indicates that the machine learning model successfully explains a substantial and highly significant proportion of the variability in the shelf-life values. Only a negligible 2.78% of the variance is left to unexplained external experimental noise. This exceptionally high percentage mathematically proves that the selected physical, chemical, and microbiological lab parameters are directly and deeply correlated with the thermodynamic and biological breakdown of the Kaju Katli matrix.

2. Mean Absolute Error (MAE)

The Mean Absolute Error measures the average absolute magnitude of the residuals-meaning the direct arithmetic distance between the AI's predicted shelf life and the actual, verified shelf life observed in the laboratory. Unlike other metrics such as Root Mean Squared Error

(RMSE), MAE treats all prediction errors linearly, preventing a single anomaly from falsely skewing the model's perceived accuracy.

The derived MAE of **0.79 days** demonstrates that across all simulated validation scenarios, the model's predictions deviate from the empirical laboratory baseline by an average of just **19 hours**. For highly perishable dairy-and-nut-based confectionery products, achieving an error window of less than one single day represents a major leap in operational precision.

4.4.4. Feature Importance Analysis

While predicting the exact remaining shelf life is crucial, understanding the mathematical and physical drivers behind the model's decisions provides significant engineering value. The Random Forest Regressor architecture provides built-in feature importance scores calculated via Mean Decrease in Impurity (Gini Importance). This analysis ranks each experimental laboratory input based on its overall contribution toward reducing prediction errors across all 100 decision trees.

The relative mathematical weights of the tracking parameters are illustrated

Free Fatty Acids (FFA %) exhibited the highest feature importance score (63.18%), suggesting that it was the most influential factor affecting shelf-life prediction in the machine learning framework. This was followed by Yeast and Mold Count (YMC) at 19.22%, while physical matrix metrics like Water Activity (a_w) scored lower at 3.80%.

There are definitive biochemical, thermodynamic, and computational reasons that explain this feature hierarchy explaining how a decreasing water activity profile maps to a simultaneous decrease in remaining shelf life:

1. The Intermediate Moisture Food (IMF) Matrix and the Water Activity Paradox

In high-moisture food systems (such as dairy milk or fresh meat), a decrease in water activity implies an increase in microbial stability and an extension of shelf life. However, Kaju Katli structurally behaves as an Intermediate Moisture Food (IMF) dominated by a dense nut-lipid and crystalline carbohydrate network.

Over the storage timeline, the product undergoes continuous desiccation, losing moisture via ambient evaporation. As a physical consequence of this total water volume loss, the measured

thermodynamic water activity (a_w) steadily decreases over consecutive days. Rather than acting as a preservation mechanism, this declining water activity trajectory serves as a physical fingerprint of stale aging, moisture migration, and terminal textural desiccation (hardening). Thus, as water activity drops, the remaining shelf life mathematically decreases because the product is physically moving closer to its unpalatable staling endpoint.

2. FFA Accumulation as the Absolute Chemical Metric of Spoiling

While a drop in (a_w) signals that the sweet is drying out, the algorithm prioritized FFA % (63.18%) because it acts as the primary chemical indicator of lipid rancidity. Kaju Katli features a high natural concentration of unsaturated fats derived from ground cashew nuts. Over time, environmental exposure drives the hydrolytic cleavage of these cashew triglycerides, leading to a steady, non-linear accumulation of free fatty acids. Because FFA levels map directly to the development of sharp off-flavors and sensory failure, the Random Forest model recognized fat acidity as a far more definitive mathematical predictor of structural expiry than the linear physical drying curve.

3. Persistent Biological Gating and Secondary Environmental Triggers

Yeast and Mold Count (YMC) achieved the second-highest importance weight (19.22%) due to the highly xerophilic (dry-tolerant) nature of fungal spoilage organisms. The machine learning model discovered that even though the water activity decreased continuously throughout the study, it remained trapped within the precise threshold zone (typically between 0.65 and 0.75) where specialized molds can actively germinate.

Because the decrease water activity was never sufficiently less to completely arrest biological activity, the algorithm treated secondary factors like Storage Condition (4.79%), Moisture (4.12%), and Water Activity (3.80%) as baseline environmental catalysts. They control the kinetic speed of degradation, but the actual, definitive progress of the sweet's decay is measured by the rising **FFA** and **YMC** profiles.

CHAPTER V

CONCLUSION

Traditional Indian confectionery products, or *mithais*, present unique challenges in modern food preservation pipelines due to their intricate, heterogeneous matrices. *Kaju katli*, a premium confectionery formulation composed primarily of ground cashew kernels and sucrose, represents an intermediate-moisture food system with specific chemical and biological vulnerabilities. The study successfully integrated multi-parameter laboratory degradation tracking with advanced ensemble machine learning techniques to establish an automated, non-destructive shelf-life predictive mode

The investigation followed a rigid, five-stage methodology balancing traditional food chemistry with modern computational processing. First, *kaju katli* samples were standardized to a commercially typical premium ratio of 65% cashew nut paste and 35% sucrose, prepared under controlled thermal processing parameters without preservatives to eliminate batch-to-batch variations. Second, the prepared samples were split and stored under two distinct environments to simulate real-world supply chain scenarios: Ambient Storage (32 ± 2°C) and Refrigerated Storage (7 ± 2°C).

Third, a rigorous longitudinal sampling schedule was carried out. Physicochemical metrics-including moisture, water activity, and free fatty acids (FFA) were compiled alongside microbiological indices. These microbiological markers included Standard Plate Count (SPC) and Yeast and Mold Count (YMC). Fourth, to overcome asynchronous data sampling frequencies and sparse data structures, missing temporal gaps were handled using linear and logarithmic interpolation, expanding the sparse observations into a dense, daily master dataset (KajuKatli_ML_Ready_Combined2.csv). Finally, categorical variables were binary-encoded, and the normalized feature matrix was processed using an 80:20 train-test split to train a cloud-hosted Ensemble Random Forest Regressor.

Under ambient conditions, shelf life terminated rapidly at Day 12 due to aggressive fungal proliferation. The YMC surged to 1.94 log CFU/g by Day 12, approaching the maximum permissible FSSAI food safety limit of 2.0 log CFU/g. This rapid colonization by osmophilic yeasts and xerophilic molds was facilitated by a relatively high ambient water activity workspace.

Conversely, controlled refrigerated storage extended product viability to Day 20, shifting the primary spoilage mechanism. In cold storage, fungal germination was heavily suppressed, but the prolonged storage timeline induced significant moisture migration and matrix desorption,

dropping moisture content to a low of 4.76% and triggering textural staling (hardening). Furthermore, terminal refrigerated spoilage was dictated by psychrotrophic bacterial overgrowth, with SPC breaching the FSSAI maximum permissible safety ceiling (4.88log CFU/g) to reach 5.26 log CFU/g at Day 20. Hydrolytic rancidity was also pronounced under refrigeration, accumulating an FFA value of 1.51% due to the continuous activity of endogenous lipolytic enzymes over the extended 20-day storage window.

The Machine Learning architecture yielded exceptionally high predictive precision when validated against unseen testing partitions. The Ensemble Random Forest model achieved a Coefficient of Determination (R^2) score of 97.22%, proving that the selected independent physical, chemical, and biological features account for nearly all variance in product degradation. The model recorded a Mean Absolute Error (MAE) of just 0.79 days, confirming that the algorithmic framework can forecast remaining shelf life to within an average window of 19 hours.

Feature importance analysis via Gini impurity reduction revealed that Free Fatty Acids (FFA %) exerted the most dominant predictive weight at 63.18%, followed by YMC at 19.22%. This proved that the random forest blocks recognized chemical rancidity and fungal counts as the absolute, defining kinetic markers of product spoilage. Conversely, physical metrics like water activity (a_w) scored lower at 3.80%. Because *kaju katli* acts as an intermediate moisture food, a declining water activity curve serves as a structural fingerprint of matrix drying and staling rather than a biological preservation mechanism.

This research bridges the gap between empirical food processing trials and computational intelligence, introducing several scalable avenues for future exploration

In conclusion, this research successfully establishes data-driven, automated soft-sensing modeling as a highly accurate alternative to destructive and costly laboratory testing in traditional food engineering sectors.

CHAPTER VI

REFERENCES

Arteaga-Cabrera, E., Ramírez-Márquez, C., Sánchez-Ramírez, E., Segovia-Hernández, J. G., Osorio-Mora, O. and Gómez-Salazar, J. A. 2025. Advancing optimization strategies in the food industry: From traditional approaches to multi-objective and technology-integrated solutions. Appl. Sci. 15(7): 3846. https://doi.org/10.3390/app15073846

Arunima, P. L., Gopinath, P. P., Lekshmi, P. G. and Esakkimuthu, M. 2024. Digital assessment of post-harvest Nendran banana for faster grading: CNN-based ripeness classification model. Postharvest Biol. Technol. 214: 112972. https://doi.org/10.1016/j.postharvbio.2024.112972

Baskar, K. V., Sri, R. N., Lavanya, S. and Harisha, V. 2024. Deep CNN model for predicting shelf life of fresh fruits and vegetables using temperature simulation data for optimized transport and storage. J. Educ. Teach. Train. 15(5): 425-434.

Bauer, A. S., Leppik, K., Galić, K., Anestopoulos, I., Panayiotidis, M. I., Agriopoulou, S., Milousi, M., Uysal-Unalan, I., Varzakas, T. and Krauter, V. 2022. Cereal and confectionary packaging: Background, application and shelf-life extension. Foods 11(5): 697. https://doi.org/10.3390/foods11050697

Beuchat, L. R. 2002. Ecological factors influencing survival and growth of human pathogens on raw fruits and vegetables. Microbes Infect. 4(4): 413-423. https://doi.org/10.1016/s1286-4579(02)01555-1

Bhargava, A. and Bansal, A. 2021. Fruits and vegetables quality evaluation using computer vision: A review. J. King Saud Univ. Comput. Inf. Sci. 33(3): 243-257. https://doi.org/10.1016/j.jksuci.2018.06.002

Cao, J., Xiong, J., Zhang, C., Zhao, Y., Guo, X., Sang, Y., Wang, H. and Zhou, Y. 2026. Development of a soil cadmium bioaccessibility prediction model for human health risk

assessment. iScience 29(6). https://doi.org/10.1016/j.isci.2026.115679

Elizabeth, J. R., Kethsial, M. S. R., Kavitha, K. and Kalpana, G. 2026. Machine learning-enhanced shelf-life prediction model for packaged foods using environmental cues. Int. J. Adv. Signal Image Sci. pp. 262-273. https://doi.org/10.29284/kttd4e71

Fellows, P. J. 2022. Food Processing Technology: Principles and Practice. Woodhead Publishing, Cambridge, UK. https://doi.org/10.1016/C2019-0-04416-0

Fennema, O. R., Damodaran, S. and Parkin, K. L. 2017. Introduction to food chemistry. In: Fennema's Food Chemistry. 5th Ed. CRC Press, Boca Raton, Florida. pp. 1-16. https://doi.org/10.1201/9781315372914-1

Gama, T., Wallace, H. M., Trueman, S. J. and Hosseini-Bai, S. 2018. Quality and shelf life of tree nuts: A review. Sci. Hortic. 242: 116-126. https://doi.org/10.1016/j.scienta.2018.07.036

Goyal, S. and Goyal, G. K. 2012. Machine learning Elman technique for predicting shelf life of burfi. Int. J. Mod. Educ. Comput. Sci. 4(7): 17. https://doi.org/10.5815/ijmecs.2012.07.03

Hosen, M. A. and Galib, S. M. 2025. Quality and shelf-life prediction of cauliflower using machine learning under vacuum and modified atmosphere packaging. Int. J. Artif. Intell. 14(2): 907-916. https://doi.org/10.11591/ijai.v14.i2.pp907-916

Jimeno, B., Rosales, A. and Domingo, C. J. 2020. Shelf-life prediction for Indian mango roll using accelerated shelf-life testing (ASLT) shelf-life plot approach. Food Res. 4: 839-845. https://doi.org/10.26656/fr.2017.4(3).399

Li, H., Li, C., Shoaib, M., Zhang, W. and Murugesan, A. 2025. Advances in non-thermal processing of meat and monitoring meat protein gels through vibrational spectroscopy. Foods 14(11): 1929. https://doi.org/10.3390/foods14111929

Lima, J. R., Bruno, L. M., Wurlitzer, N. J., Sousa, P. H. M. and Holanda, S. A. M. 2021. Cashew nut-based beverage: Development, characteristics and stability during refrigerated storage.

Food Sci. Technol. 41: 60-64. https://doi.org/10.1590/fst.33819

Lokhande, S. P., Waseem, M., Datir, R. P., Dhotre, A. V. and Wasnik, P. G. 2020. Storage studies on textural aspects of selected Indian dairy products. Indian J. Dairy Sci. 73(3): 216-221. https://doi.org/10.33785/IJDS.2020.v73i03.004

Patel, A. M., Momin, J. K. and Damle, K. S. 2021. Development of sugar-free Kajukatli: A traditional Indian sweet. Int. J. Chem. Stud. 9(4): 238-241.

Raisi, M., Ghorbani, M., Mahoonak, A. S., Kashaninejad, M. and Hosseini, H. 2015. Effect of storage atmosphere and temperature on the oxidative stability of almond kernels during long term storage. J. Stored Prod. Res. 62: 16-21. https://doi.org/10.1016/j.jspr.2015.03.004

Rico, R., Bulló, M. and Salas-Salvadó, J. 2016. Nutritional composition of raw fresh cashew (Anacardium occidentale L.) kernels from different origin. Food Sci. Nutr. 4(2): 329-338. https://doi.org/10.1002/fsn3.294

Sharma, A. K., Brahmbhatt, J. V. and Patel, A. M. 2017. Storage study of standardized Kajukatli. Int. J. Sci. Res. Technol. 3(5): 406-409.

Sonwani, E., Bansal, U., Alroobaea, R., Baqasah, A. M. and Hedabou, M. 2022. An artificial intelligence approach toward food spoilage detection and analysis. Front. Public Health 9: 816226. https://doi.org/10.3389/fpubh.2021.816226

Tarlak, F. 2023. The use of predictive microbiology for the prediction of the shelf life of food products. Foods 12(24): 4461. https://doi.org/10.3390/foods12244461

Trox, J., Vadivel, V., Vetter, W., Stuetz, W., Scherbaum, V., Gola, U., Nohr, D. and Biesalski, H. K. 2010. Bioactive compounds in cashew nut (Anacardium occidentale L.) kernels: Effect of different shelling methods. J. Agric. Food Chem. 58(9): 5341-5346. https://doi.org/10.1021/jf904580k

Tupe, R. and Ananthanarayan, L. 2020. Estimation of starch and sugars for detection of potential

adulteration in kaju katli, a popular cashew nut based Indian confection. J. Microbiol. Biotechnol. Food Sci. 9(6): 1151–1156. https://doi.org/10.15414/jmbfs.2020.9.6.1151-1156

Wunderlich, P., Pauli, D., Neumaier, M., Wisser, S., Danneel, H. J., Lohweg, V. and Dörksen, H. 2023. Enhancing shelf-life prediction of fresh pizza with regression models and low cost sensors. Foods 12(6): 1347. https://doi.org/10.3390/foods12061347

APPENDIX I

1. Raw dataset obtained from data collection for shelf-life study in case of ambient conditions

Date	Water Activity	Moisture Content (%)	FFA (%)	SPC (log cfu/g)	YMC (log cfu/g)
10-Mar	0.72	6.86	1.32		
		6.77	1.34		
		6.83	1.33		
11-Mar	0.717			initial (0)	initial (0)
12-Mar	0.712				
13-Mar	0.709	6.14	1.35	2.46	
		6.23	1.37		
		6.11	1.34		
14-Mar	0.705			3.07	1.51
15-Mar	0.703				
16-Mar	0.698	5.88	1.39	3.65	
		5.81	1.38		
		5.89	1.41		
17-Mar	0.694			4.14	1.76
18-Mar	0.687				
19-Mar	0.683	5.73	1.41	4.26	
		5.69	1.4		
		5.65	1.38		
20-Mar	0.677				1.94
21-Mar	0.675			4.39	
22-Mar	0.673				
23-Mar	0.671	5.35	1.44	4.48	
		5.24	1.46		
		5.27	1.47		
24-Mar	0.668			4.54	2.62

2. Raw dataset obtained from data collection for shelf-life study in case of refrigerated conditions

Date	Water Activity	Moisture Content (%)	FFA (%)	SPC	YMC
10-Mar	0.721	6.86	1.32		
		6.77	1.34		
		6.83	1.33		
11-Mar	0.72			initiation (0)	0
12-Mar	0.718				
13-Mar	0.717	6.65	1.34	2.42	
		6.62	1.35		
		6.71	1.33		
14-Mar	0.715			2.86	
15-Mar	0.714				
16-Mar	0.712	5.96	1.36	3.27	1.24
		6.34	1.38		

		6.18	1.39		
17-Mar	0.709			3.67	
18-Mar	0.704				
19-Mar	0.7	5.92	1.41	3.84	
		5.84	1.42		
		5.77	1.42		
20-Mar	0.698				
21-Mar	0.692			4.24	1.43
22-Mar	0.687				
23-Mar	0.683	5.35	1.46	4.31	
		5.23	1.47		
		5.41	1.49		
24-Mar	0.679				
25-Mar	0.676			4.86	
26-Mar	0.671	5.12	1.48		1.61
		5.08	1.51		
		4.98	1.5		
27-Mar	0.668			5.12	
28-Mar	0.664				
29-Mar	0.659	4.73	1.51	5.26	
		4.82	1.52		
		4.69	1.49		
30-Mar	0.656				1.79
31-Mar	0.652			5.38	
1-Apr	0.647	4.2	1.53		
		4.13	1.55		
		4.09	1.57		
2-Apr	0.646				
3-Apr	0.643				
4-Apr	0.642	3.97	1.54	5.49	1.92
		4.06	1.58		
		4.02	1.59		
5-Apr	0.641				

3. Combined dataset obtained after interpolation and feature engineering used for training shelf-life prediction model

0Ambient	0	0.72	6.82	1.33	0	0	11
1Ambient	0	0.717	6.6	1.337	1.23	0.378	10
2Ambient	0	0.712	6.38	1.343	2.46	0.755	9
3Ambient	0	0.709	6.16	1.35	2.765	1.133	8
4Ambient	0	0.705	6.06	1.363	3.07	1.51	7
5Ambient	0	0.703	5.96	1.377	3.36	1.593	6
6Ambient	0	0.698	5.86	1.39	3.65	1.677	5
7Ambient	0	0.694	5.803	1.397	4.14	1.76	4
8Ambient	0	0.687	5.747	1.403	4.2	1.82	3
9Ambient	0	0.683	5.69	1.41	4.26	1.88	2
10Ambient	0	0.677	5.583	1.427	4.325	1.94	1
11Ambient	0	0.675	5.477	1.443	4.39	2.11	0

0Refrigerated	1	0.721	6.86	1.32	0	0		20
1Refrigerated	1	0.72	6.79	1.327	1.21	0.207		19
2Refrigerated	1	0.718	6.72	1.333	2.42	0.413		18
3Refrigerated	1	0.717	6.65	1.34	2.64	0.62		17
4Refrigerated	1	0.715	6.547	1.347	2.86	0.827		16
5Refrigerated	1	0.714	6.443	1.353	3.065	1.033		15
6Refrigerated	1	0.712	6.34	1.36	3.27	1.24		14
7Refrigerated	1	0.709	6.2	1.377	3.47	1.278		13
8Refrigerated	1	0.704	6.06	1.393	3.67	1.316		12
9Refrigerated	1	0.7	5.92	1.41	3.755	1.354		11
10Refrigerated	1	0.698	5.777	1.422	3.84	1.392		10
11Refrigerated	1	0.692	5.635	1.435	4.04	1.43		9
12Refrigerated	1	0.687	5.492	1.448	4.24	1.466		8
13Refrigerated	1	0.683	5.35	1.46	4.275	1.502		7
14Refrigerated	1	0.679	5.26	1.467	4.31	1.538		6
15Refrigerated	1	0.676	5.17	1.473	4.585	1.574		5
16Refrigerated	1	0.671	5.08	1.48	4.86	1.61		4
17Refrigerated	1	0.668	4.963	1.49	4.99	1.646		3
18Refrigerated	1	0.664	4.847	1.5	5.12	1.682		2
19Refrigerated	1	0.659	4.73	1.51	5.19	1.718		1
20Refrigerated	1	0.656	4.553	1.517	5.26	1.754	0	

APPENDIX II

1. **Code for deployment and execution of ensemble Random Forest Model**

i. Importing the dataset using pandas library

```
import pandas as pd
import numpy as np

# 1. Load the dataset
# If your filename is different, change it inside the quotes
below
df = pd.read_csv("KajuKatli_Data.csv")

# 2. Drop completely empty rows that have no data
df = df.dropna(how='all')

# 3. Drop rows where the target 'RemainingShelfLife' is missing
df = df.dropna(subset=['RemainingShelfLife'])

# 4. View the structure and details of your clean dataset
Print"--- Data Info ---")
print (df.info ())
print"\n--- First 5 Rows ---")
print(df.head())
```

ii. Scikit learn used to split the dataset into training data (26 rows) and testing data (7 rows)

```
from sklearn.model_selection import train_test_split

# 1. Select your input features (X) and remove text/target
columns
X = df.drop(columns='Storage', 'RemainingShelfLife'])

# 2. Select your target column (y)
y = df['RemainingShelfLife']

# 3. Split the data: 80% for training the AI, 20% for testing
its accuracy
# Notice the change below: 'test_size' is used instead of
'test_test_split'
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

print (f"Training data size: {X_train.shape[0]} rows")
print (f"Testing data size: {X_test.shape[0]} rows")
```

iii. Importing and implementation of Random Forest Regressor

```
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_absolute_error, r2_score

# 1. Initialize the Random Forest model
```

```
model = RandomForestRegressor(n_estimators=100, random_state=42)

# 2. Train the model using your training data
model.fit(X_train, y_train)

# 3. Test the model using your test data
predictions = model.predict(X_test)

# 4. Evaluate how accurate the model is
mae = mean_absolute_error(y_test, predictions)
r2 = r2_score(y_test, predictions)

print ("--- Model Performance ---")
print (f"Mean Absolute Error (MAE): {mae:.2f} days")
print (f"R-squared Score (Accuracy): {r2 * 100:.2f}%")
```

iv. Testing the model based on random data

```
import pandas as pd

# Update these numbers based on your latest lab test results
new_batch = pd.DataFrame([{
    'Day': 3,                # Days passed since production
    'Storage_Code': 1,         # 0 for Ambient (Room Temp), 1 for
Refrigerated
    'WaterActivity': 0.715,    # Lab Water Activity result
    'Moisture': 6.54,          # Lab Moisture %
    'FFA': 1.34,               # Free Fatty Acids value
    'SPC': 2.64,               # Standard Plate Count value
    'YMC': 0.62                # Yeast and Mold Count value
}])

# Get the prediction from your trained AI
predicted_shelf_life = model.predict(new_batch)
print (f"Predicted Remaining Shelf Life: {predicted_shelf_life[0]:.1f}
days")
```

5)

DEVELOPMENT OF A MACHINE LEARNING BASED SHELF-LIFE PREDICTION MODEL FOR KAJU KATLI

by

ALEENA BIJU (2022-02-047)

SOURAV P (2022-02-050)

NAVYA ANKITA ANAND (2022-02-051)

PROJECT REPORT

Submitted in partial fulfillment of the requirement for the degree of

BACHELOR OF TECHNOLOGY

in

AGRICULTURAL ENGINEERING

Faculty of Agricultural Engineering and Technology

Kerala Agricultural University

DEPARTMENT OF PROCESSING AND FOOD ENGINEERING

KELAPPAJI COLLEGE OF AGRICULTURAL ENGINEERING AND FOOD TECHNOLOGY, TAVANUR-679573,

MALAPPURAM, KERALA, INDIA

2026

ABSTRACT

Kaju katli is a widely consumed traditional Indian confectionery prepared primarily from cashew nuts and sugar. Despite its popularity, the product has a limited shelf life due to physicochemical and microbiological deterioration during storage, including moisture loss, lipid oxidation, and microbial growth. Conventional shelf-life determination methods rely on repeated laboratory testing, which is time-consuming, labour-intensive, and expensive. This study aims to develop a machine learning-based approach for predicting the shelf life of kaju katli using quality parameters obtained through laboratory analysis. The work focuses on integrating food science with artificial intelligence to provide a rapid, reliable, and cost-effective alternative to conventional shelf-life estimation.

The study evaluates key physicochemical and microbiological parameters, namely moisture content, water activity (aw), free fatty acid (FFA), standard plate count (SPC), and yeast and mould count (YMC), as indicators of product quality during storage. Experimental data collected from laboratory analyses are organized and enhanced through feature engineering techniques to improve dataset quality and predictive capability. A Random Forest regression model is developed to establish relationships between these quality attributes and the remaining shelf life of kaju katli. The model is trained and evaluated using standard performance metrics to ensure prediction accuracy and reliability.

The proposed machine learning model is expected to accurately predict shelf life by capturing the complex interactions among physicochemical and microbiological parameters that influence product deterioration. The approach minimizes dependence on prolonged storage studies while providing a scientific basis for quality monitoring and decision-making. The findings are expected to demonstrate that moisture content, water activity, FFA, SPC, and YMC are significant predictors of shelf life and can effectively support intelligent food quality assessment.

Overall, this research highlights the potential of machine learning in modern food engineering by offering an efficient tool for shelf-life prediction of traditional confectionery products. The developed model can help manufacturers optimize storage conditions, reduce food wastage, improve inventory management, and enhance consumer confidence by providing more accurate shelf-life estimates. Furthermore, the methodology can be extended to other traditional Indian sweets and similar food products, contributing to the advancement of intelligent food processing and quality assurance systems.