

Machine Learning Algorithm for Breastmilk Quality Classification Using Multi-Array Sensor Technology: A Systematic Literature Review

M. Shahkhir Mozamir*, Iqlima Nadhira Jamaludin, Shafina Binti Abd Karim Ishigaki, Fatin Aliah Binti Yahya, Anggi Muhammad Rifa'i

Faculty Technology Maklumat dan Komunikasi, University Technical Malaysia Melaka, Malaysia

DOI: <https://dx.doi.org/10.47772/IJRISS.2025.909000797>

Received: 29 September 2025; Accepted: 04 October 2025; Published: 30 October 2025

ABSTRACT

Assessing breastmilk quality is essential to ensure optimal nutrition for infants. However, conventional laboratory-based methods are often time-consuming, costly, and impractical for real-time applications. Recent advancements in multi-array sensor technology combined with machine learning algorithms present a promising solution for efficient and accurate breastmilk classification. This systematic literature review aims to evaluate existing research on the integration of sensor technologies and machine learning models for breastmilk quality assessment. It specifically addresses the shortcomings of traditional approaches and explores the feasibility of real-time monitoring systems. Following the PRISMA guidelines, research articles were initially collected at 42 papers published between 2020 and 2025. The papers identified from IEEE Xplore, Scopus, ScienceDirect, and Google Scholar databases. After applying inclusion criteria and a thorough screening process, unfortunately only 5 research papers were selected based on their relevance to sensor integration, machine learning algorithms, dataset characteristics, and classification performance. Applying machine learning for breastmilk classification is arising by year. The findings categorize existing approaches into traditional statistical methods, machine learning techniques, and deep learning models. Random Forest and Support Vector Machines (SVM) emerged as commonly used classifiers due to their balance between accuracy and computational efficiency. Although deep learning models show potential for improved accuracy, they require larger datasets and greater processing power. Feature extraction and selection significantly influence classification outcomes by identifying key breastmilk components. This review provides a foundation for developing real-time breastmilk quality monitoring systems using multi-array sensors and machine learning, offering valuable insights for advancing maternal and infant healthcare technologies.

Keywords— Breastmilk Quality Classification, Multi-Array Sensor Technology, Machine Learning Algorithms, Random Forest Classifier, Real-Time Monitoring

INTRODUCTION

Breast milk is widely known as the best source of nutrition for babies. It not only provides essential nutrients but also contains immune-boosting properties and bioactive compounds that are important for healthy growth and development [1]. However, storing and handling expressed breast milk is not always easy. One of the biggest challenges of mothers' faces is making sure the milk is still fresh and safe to feed their babies. Over time, stored breast milk can go through changes both chemical and microbial that may reduce its nutritional value or make it unsafe. Research suggests that breast milk is best used within six months of storage, but beyond that, there's a risk of spoilage. Unfortunately, current methods for keeping track of milk freshness, like writing labels on bottles, are prone to mistakes [2].

This is where technology can help. In recent years, multi-array sensors and artificial intelligence (AI) have opened new possibilities for monitoring food quality and these tools can be applied to breast milk too. Multi-

array sensors, which can include things like pH sensors, gas sensors (also known as electronic noses), and biosensors, are already used in the food industry to detect spoilage by measuring things like acidity, gases, and bacteria [3]. When combined with AI, these sensors can become part of a smart system that classifies milk freshness automatically based on the data they collect taking human error out of the picture [4].

Among the various AI tools available, the Random Forest algorithm stands out. It is reliable, accurate, and has already proven itself in many food-related applications, especially in dairy. Random Forest works by combining the results of many decision trees to make a strong prediction, for example, whether milk is fresh or spoiled. That makes it a promising choice for this kind of project [5].

This review sets out to explore how researchers have approached the challenge of monitoring the quality of stored breast milk using sensor technologies. It looks at the types of sensors that have been used, what indicators of spoilage they detect, and how the information gathered from these sensors is analyzed to assess milk freshness and safety. By bringing together and examining the findings from previous studies, this review aims to provide a clear picture of what has been done so far, highlight any gaps in the current research, and suggest possible directions for future work in making breast milk monitoring more accurate and reliable.

Motivation & Related Work

Limited yet growing awareness of breastmilk quality monitoring, especially using intelligent systems, has inspired the researchers to explore the integration of machine learning algorithms and multi-array sensor technologies for this purpose. While numerous studies have investigated breast milk composition and safety through conventional laboratory methods, there is a noticeable lack of research focused on automated classification approaches such as Random Forest algorithm in combination with diverse sensor arrays such as pH, molecular, and gas sensors. This research seeks to address this gap by investigating the feasibility of a smart classification system that can accurately determine the quality of stored breastmilk. By doing so, it contributes to the emerging intersection of maternal health, sensor technology, and artificial intelligence, an area that remains underexplored and full of potential.

The increasing awareness of infant health and nutrition has underscored the critical need for effective monitoring and quality assessment of breast milk. Breast milk is not only a vital source of nutrition but also a potential medium for exposure to harmful substances such as environmental contaminants and bacterial pathogens [6]. This has motivated researchers to seek innovative methods for rapid, accurate, and accessible breast milk analysis. Conventional laboratory techniques, while effective, often require expensive equipment and lengthy procedures, making them less suitable for widespread or real-time use. In response, studies have begun to explore point-of-care (POC) technologies, wearable sensors, and intelligent systems that offer practical alternatives for on-site monitoring and early detection of quality concerns [7].

Recent research has also focused on leveraging artificial intelligence and sensor-based technologies to improve both the efficiency and precision of breast milk assessment. For example, [8] advocate for the integration of wearable sensors with POC systems to enhance sensitivity and reduce contamination risks, while [9] stress the need for longitudinal studies on endocrine-disrupting chemicals (EDCs) and their effects on infant neurodevelopment. Similarly, [10] developed a portable device using multi-spectral sensors and machine learning, demonstrating its capability in rapidly determining milk composition with high accuracy. Meanwhile, [11] showcase the growing potential of AI applications for predicting breastfeeding patterns, assessing milk quality, and providing digital support to mothers pointing toward a future of more personalized and data-driven maternal care.

These collective efforts from the literature not only highlight the importance of improving breast milk monitoring technologies but also reveal current research gaps. There is a growing need for smart, integrated solutions that combine sensor hardware, AI algorithms, and real-time data analysis to enable rapid, non-invasive, and accurate assessment methods. Future research should therefore prioritize the development of robust, low-cost, and user-friendly systems that can be deployed in both clinical and home environments.

TABLE I: Summary of Motivation and Related Works

Study References	Study Focus	Contribution	Future Works
(Janakiraman et al., 2025)	●The study focuses on analyzing breast milk's contents, including nutrients and biomarkers, for better healthcare solutions for mothers and newborns. ●It discusses the advantages and drawbacks of conventional and point-of-care techniques for analyzing breast milk. ●The research highlights the increasing popularity of point-of-care analysis to combat increasing procedures due to their exposure to pollutants and portability affordability. ●Future perspectives include the integration of wearable sensors with point-of-care techniques to enhance sensitivity and reduce contamination.	●The authors contributed to the writing of the original draft, methodology, formal analysis, data curation, and conceptualization of the research. ●They also engaged in review editing, supervision, project administration, and funding acquisition. ●The paper emphasizes the need for developing more point-of-care (POC) devices for breast milk contamination analysis to combat increasing exposure to pollutants and pathogens. ●The review covers various point-of-care techniques and devices developed for detecting chemical contaminants and biomarkers in breast milk.	●Machine learning and AI systems should be incorporated into wearable sensors for better data management. ●Future work should focus on improving the accuracy and efficiency of wearable sensors. ●There is a need for more devices to detect biomarkers in breast milk. ●Smart sensors must be developed to set standards for technology integration in healthcare.
(Brambilla et al., 2025)	●The paper highlights breast milk as a key matrix for monitoring EDC exposure in infants. ●It emphasizes the need for longitudinal studies to understand EDC effects on neurodevelopment. ●The research calls for improved assessment tools to reduce confounding factors in studies. ●It advocates for public health efforts to minimize maternal and infant EDC exposure. ●The study discusses the adverse effects of EDCs on various developmental outcomes in infants.	●The study focuses on the effects of endocrine disrupting chemicals (EDCs) in breast milk on neurobehavioral development in infants. ●It evaluates the association between EDC exposure and various developmental outcomes during the first six years of life. ●The research emphasizes the need for longitudinal studies to understand EDC effects on infant neurodevelopment.	●Future work should focus on longitudinal cohort studies to monitor EDC exposure during pre- and post-natal periods. ●It is essential to assess the impact of EDCs on human breast milk and infant development. ●Developing risk models that include individual and contextual factors is necessary. ●Promoting monitoring plans for EDC exposure and public policies is crucial for infant neurodevelopment. ●The Life-MILCH Project exemplifies a longitudinal study linking EDC levels in breast milk to maternal habits and infant outcomes. ●Further longitudinal studies are needed to understand protective

			factors against EDC exposure.
(Wang et al., 2023)	●A portable detection instrument was developed to rapidly detect milk components, addressing traditional detection challenges. ●The instrument combines multi-spectral sensors, machine learning algorithms, and an embedded system for efficiency. ●A broadband NIR LED constant-current driver circuit was designed to obtain six NIR features of milk samples. ●The XGBoost model was selected for training, demonstrating better generalization for unknown milk samples. ●The proposed instrument allows for accurate measurement of protein and fat contents in a short time.	●The study focuses on developing a portable detection instrument for rapid milk composition analysis using multi-spectral sensors and machine learning algorithms. ●It addresses challenges like long measurement periods, high costs, and environmental pollution in traditional milk detection methods. ●The research evaluates the performance of the XGBoost model for predicting protein and fat content in milk samples.	●Future research will consider adding different temperatures as new features for machine model training to mitigate temperature effects on experiments.
(Agudelo-Pérez et al., 2024)	●The study identifies AI's potential to improve breastfeeding rates by analyzing human milk and targeting high-risk populations and providing tailored support. ●AI models accurately predict macronutrient content in human milk, enhancing operational efficiency for milk banks. ●AI-driven chatbots effectively address breastfeeding concerns and connect mothers to support programs, showing high engagement. ●The research highlights AI's role in detecting environmental contaminants in milk.	●The study focuses on the use of artificial intelligence (AI) in analyzing human milk and breastfeeding practices. ●It includes studies on AI applications for breastfeeding and human milk analysis. ●The research emphasizes AI's role in predicting breastfeeding patterns and analyzing milk composition. ●It also explores AI-driven educational tools for breastfeeding support.	●Not addressed in the paper.

	aiding risk assessment. ●The paper emphasizes AI's ability to analyze complex data, offering insights beyond traditional methods.		
--	--	--	--

METHODOLOGY

To guide the implementation of this study, experimental methodology was adopted and structured into three main phases: 1) Preparation, 2) Implementation, and 3) Analysis and Discussion. In the Preparation phases, the research problem was identified, and the objectives were clearly defined, focusing on the need to classify breastmilk quality using sensor-based data and machine learning. Figure 1 shows the methodology to conduct this literature review.

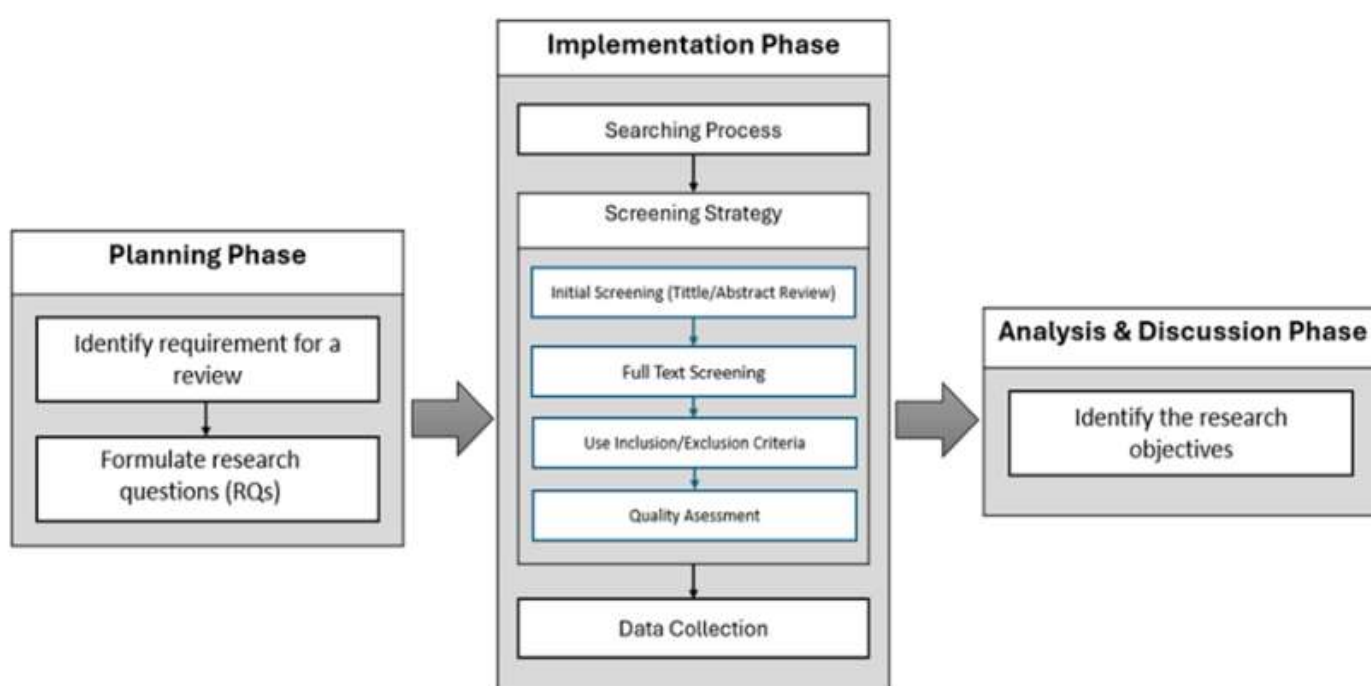


Fig. 1 Systematic Literature Review (SLR) Methodology

Planning Phase

The previous section provided an overview of the motivation and background concerning breast milk quality classification, particularly through the integration of machine learning techniques and multi-array sensor technologies. While existing literature has offered valuable insights into the use of sensors and machine learning for quality analysis, a gap remains in the focused exploration of specific algorithms, particularly in the context of detecting breast milk quality based on pH levels, composition, and aroma. Many reviews in this field overlook the structural design, implementation stages, and performance of these algorithms, especially in tasks tailored to the unique properties of breastmilk. This study aims to address this gap by conducting an SLR that provides a structured evaluation of the algorithms applied in breastmilk quality classification. The goal of this SLR is to thoroughly assess the methodological limitations, procedural workflows, and advantages of the algorithms, with a particular emphasis on how they handle the complexities of breastmilk analysis. To ensure that the review remains closely aligned with its research objectives, a set of Research Questions (RQs) was formulated. These questions will guide the process of selecting, analyzing, and synthesizing relevant studies. The rationale behind each research question and its relevance to the broader aims of the review will be presented in Table 2.

TABLE II: Research Questions (RQs)

Research Questions (RQs)	Motivations
RQ1: What types of papers are covered by the investigation?	To identify the different sets of findings in the domain.
RQ2: Which are the most commonly used or compared algorithms of classification used in breastmilk classification?	To determine the classification algorithms that are frequently applied or evaluated in breastmilk classification.
RQ3: What types and how many sensors are typically used in breastmilk classification?	To identify the sensor technologies and quantity used in multi-sensor arrays to classify the breastmilk quality.
RQ4: What are the performances metrics used to evaluate classification algorithms in breastmilk classification?	To identify how the performance of classification models is measured, including their accuracy, reliability and overall effectiveness in breastmilk classification.
RQ5: What kind of field research filed that is study in breastmilk classification?	To understand the key focus areas in breastmilk related research field.
RQ6: What are the key challenges and limitations associated with classification algorithms in breastmilk classification?	To determine the challenges and limitations for the odour classification.
RQ7: What are the future trends and potential research directions for classification algorithms in breastmilk classification?	To find new trends, directions for further research, and ideas in the context of breastmilk classification

Implementation Phase

The implementation phase will focus on applying the selected machine learning algorithms for breastmilk quality classification using multi-array sensor data, including pH, composition, and aroma. After preprocessing the data, the algorithms will be integrated with sensor technologies, followed by training and testing to assess their performance. Validation will be conducted using cross-validation and performance metrics such as accuracy and precision to ensure the models' effectiveness in classifying breastmilk quality. The process will be iterative, refining the models based on their performance and addressing domain-specific challenges.

Searching Process

This phase involves executing the defined stages including a systematic approach for identifying relevant studies, method used for selecting the articles to be included and the process of data collection and analysis. A well-defined search process is crucial for achieving high-quality and reliable results. In this study, the identification and selection of sources will be carried out systematically to gather all relevant studies related to the topic. This process is guided by two key elements: 1) Use various search strings. 2) Selecting appropriate electronic library databases for resources.

The search terms are formulated based on the Research Questions (RQs) and follow a standard procedure, which includes the following steps:

1. Identifying keywords related to the Research Questions (RQs).
2. Including synonyms and alternative spellings for the terms.
3. Verifying the relevance and appropriateness of the selected search terms.

4. Combining the search terms using Boolean operators such as OR and AND.

The result for the search strings after following the steps above as below:

1. ("Breast Milk Quality Classification" OR "Human Milk Quality Classification") AND ("Multi-Array Sensor" OR "Sensor Array") AND ("Machine Learning" OR "Classification Algorithm")
2. ("Breast Milk Freshness Detection" OR "Human Milk Spoilage Detection") AND ("Multi-Sensor Technology" OR "Electronic Nose" OR "Biosensors")
3. ("Milk Quality Monitoring" OR "Milk Spoilage Detection") AND ("Multi-Array Sensor" OR "Gas Sensors" OR "Optical Sensors") AND ("Machine Learning" OR "breastmilk quality" AND "sensor array")
4. "Classification" AND "Machine Learning" AND "Breast Milk"

The second element is the selection of resources. In this systematic literature review (SLR), three electronic databases were used for the search process: Scopus, IEEE Xplore, and ScienceDirect. These databases are well-established resources containing empirical studies and literature surveys in areas of sensor technology, machine learning and healthcare analysis.

Screening Strategy

An essential aspect of this systematic literature review (SLR) is the development of a comprehensive screening strategy to ensure the inclusion of only the most relevant and high-quality studies. Following the initial search, a total of 42 articles were retrieved across the selected databases. To narrow down this collection and focus on studies directly aligned with the research objectives, a two-stage selection process was employed:

1. Applying inclusion and exclusion criteria
2. Assessing the studies based on predefined Quality Standard Questions (QSQ).

TABLE III: Inclusion and Exclusion Search Standard

Inclusion Search Standard	Exclusion Search Standard
Studies must be written in the English language.	Studies that are not written in English language
Studies have potential to answer the Research Questions (RQs) based on keywords, title and abstract.	Studies will avoid duplicating the copies, review paper and only the complete version included for this SLR.
Studies focusing on breastmilk classification, classification using sensor-based.	Studies that do not focus on breastmilk classification or use methods unrelated to sensors or classification.
Studies must highlight challenges, methods or future developments related to breastmilk quality classification.	Studies that do not address technical or methodological issues in breastmilk quality classification.

The first step involves the application of inclusion and exclusion criteria to filter out studies that do not meet the requirements of the review. Only studies that specifically focus on the research questions and align with the scope of the systematic review are retained. In the second stage, each study is assessed based on its quality. A structured evaluation using Quality Standard Questions (QSQ) ensures that the selected articles are methodologically sound and relevant to the research objectives. Each study is scored on a three-point scale, where 3 marks is for the study fully meets the criterion, 2 marks is for the partially criterion and 1 mark is for not meet the criterion.

TABLE IV: Quality Standard Question (QSQ)

QSQ ID	Inclusion Search Standard	Exclusion Search Standard
QSQ1	Are the aims of studies clearly stated?	Yes= 3 / moderate = 2 / no = 1
QSQ2	Are the context of studies well defined?	Yes= 3 / moderate = 2 / no = 1
QSQ3	Does the study focus on RQ in the specified domain?	Yes= 3 / moderate = 2 / no = 1
QSQ4	Are the proposed algorithms in studies well-explained?	Yes= 3 / moderate = 2 / no = 1
QSQ5	Is the proposed algorithm improving the localization accuracy compared to the comparison algorithm?	Yes= 3 / moderate = 2 / no = 1
QSQ6	Is the proposed algorithm improving the time computation compared to the comparison algorithm?	Yes= 3 / moderate = 2 / no = 1
QSQ7	Is the result well explained?	Yes= 3 / moderate = 2 / no = 1

TABLE V: Quality Standard Question (QSQ) scores of the 17 of studies

Reference	QSQ1	QSQ2	QSQ3	QSQ4	QSQ5	QSQ6	QSQ7	Score
[12]	3	3	2	3	3	1	3	18
[13]	3	3	2	1	2	1	2	14
[14]	3	2	1	1	2	1	2	12
[15]	3	2	3	1	1	1	3	14
[16]	3	2	1	2	1	1	2	13
[17]	3	2	1	1	1	1	2	11
[18]	2	1	1	3	1	2	3	13
[19]	3	1	1	2	3	1	3	13
[20]	3	3	1	1	1	1	2	12
[21]	3	1	1	1	1	1	3	11
[22]	1	2	1	3	2	1	3	13
[23]	3	3	2	2	1	1	2	14
[24]	3	1	1	1	1	1	2	10
[25]	3	2	1	1	1	1	2	11
[26]	3	3	3	2	1	1	1	14
[22]	1	2	2	3	3	1	1	13
[27]	3	2	1	1	2	1	3	13

Based on the assessment outlined in Table 4, the authors collaboratively reviewed and discussed the findings of each selected study to resolve discrepancies and ensure consistency in evaluation. To maintain the reliability and integrity of the review results, studies that received a quality score below 14 (less than half of the maximum score of 21) were excluded from this SLR. Following this quality screening process, a total of 5 studies were deemed suitable to serve as the primary research sources for this review. The outcome of the screening strategy is illustrated in Figure 2.

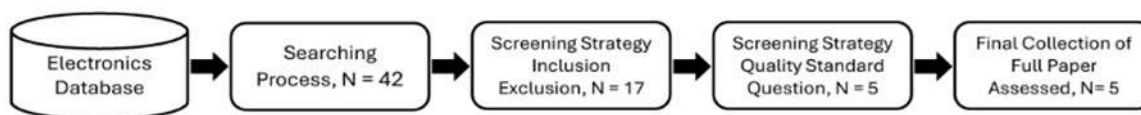


Fig. 2 Screening Strategy

RESULT AND DISCUSSION

Distribution of Studies (RQ1)

Final studies were selected for final screening in this Systematic Literature Review are 5 papers. The sourced from three major electronic databases which are Scopus, IEEE Xplore, ScienceDirect and others. These final selections include 3 conference papers and 2 journal articles, representing 60% and 40% of the total, respectively. Figure 3 presents the distribution of publication types among the studies selected in the final screening stage.

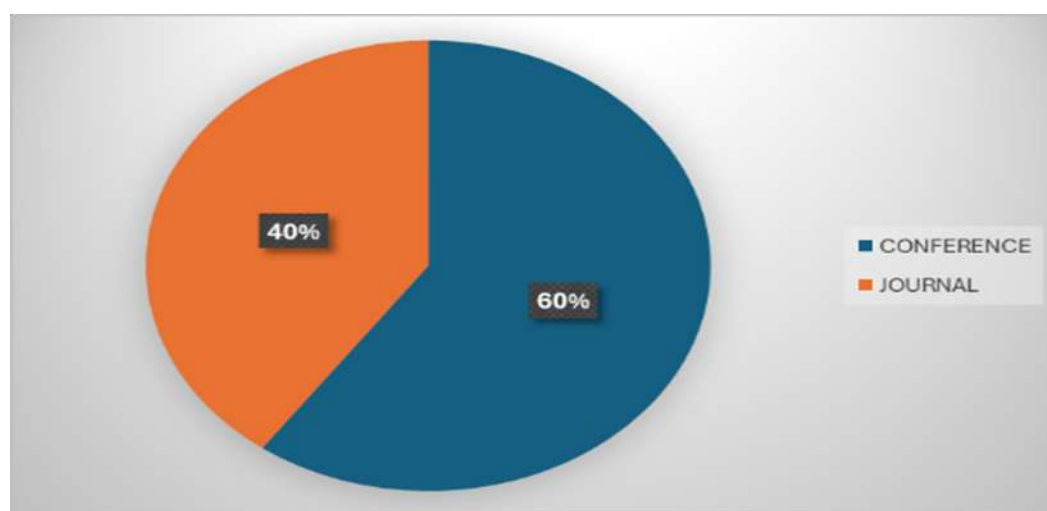


Fig. 3 Percentage of the inclusion rates by Electronic Library

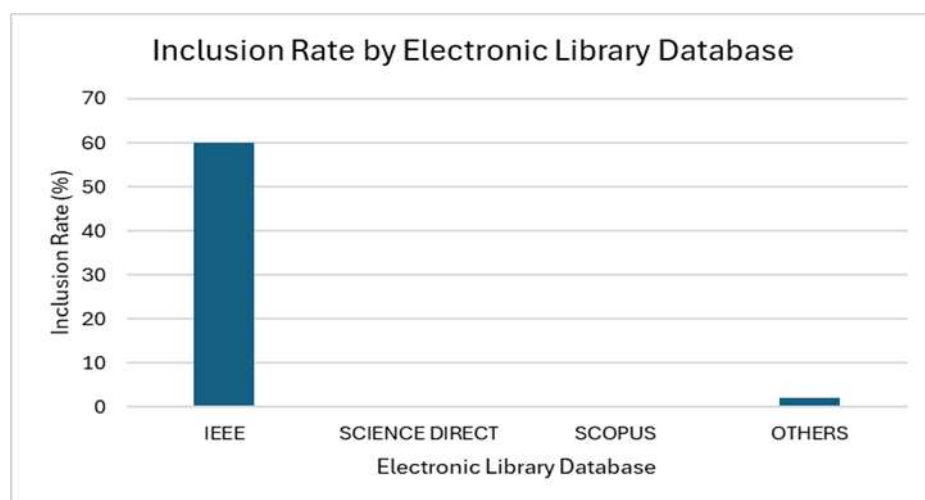


Fig. 4 Percentage of the inclusion rates by Electronic Library

TABLE VI: Number of Studies after Screening Strategies

Electronic Library	Screening Process	Screening Strategy Exclusion Including	Screening Strategy Quality Standard Question (QSQ)	
IEEE	16	7	9	3
				4

Science Direct	5	1	4	0	1
Scopus	4	1	3	0	1
Google Scholar	17	8	11	2	6
Total	42	17	24	5	12

Table 6 displays the results of the SLR screening process. Initially, 42 studies were identified across four sources: IEEE, ScienceDirect, Scopus, and Others. Through the "Screening Strategy Exclusion and Including" process, 17 studies were included and 24 were excluded. Following this, the "Screening Strategy Quality Standard Question (QSQ)" process further refined the selection, resulting in 5 included studies and 12 exclusions. Figure 4 illustrates the percentage distribution of the final included studies, with IEEE contributing approximately 60%, Scopus 20%, and others 20%. Figure 5 presents the number of final collected studies by year from 2020 to 2025, showing a general increase in publications, indicating that the research interest in the related domain remains strong and continues to grow.

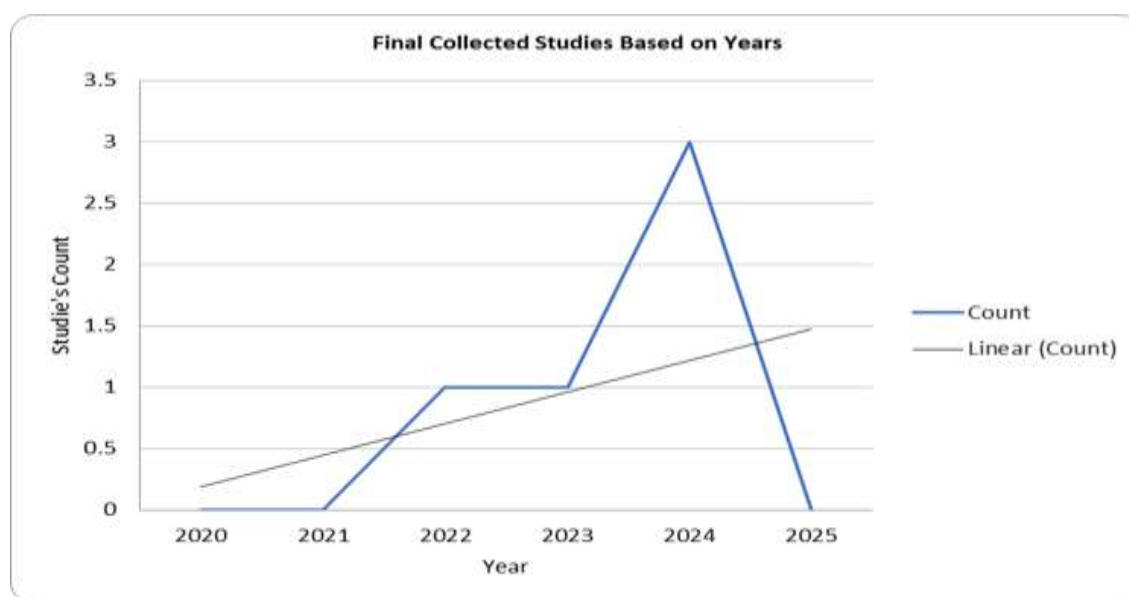


Fig. 5 Count of Studies Over Year

Which are the most used algorithms of classification used in breastmilk classification? (RQ2)

Based on the final selection of 5 reviewed studies, presents the distribution of algorithms employed in various breast milk-related applications, including ethnicity prediction based on breast milk microbiota, temperature-based alert systems during milk transportation, and chemometric analysis for breast milk quality calibration. The analysis includes both machine learning and algorithmic logic approaches tailored for biological data, transport monitoring, and spectroscopic calibration.

Each reviewed study applies a distinct algorithm tailored to its specific research objective. For example, Random Forest (RF) was employed in one study for its ensemble-based architecture, which performed well in handling noisy and high-dimensional microbiological datasets. The study highlighted RF's ability to achieve high accuracy, precision, recall, and F1-scores, demonstrating its reliability in healthcare and microbiota-based classification tasks.

In another study, Support Vector Machine (SVM) was chosen due to its effectiveness in modeling complex, non-linear relationships, particularly useful for datasets involving intricate microbial or spectral characteristics. Similarly, AdaBoost and Back Propagation Neural Network (BPNN) were applied in individual studies, mainly to evaluate classification performance or to benchmark against other machine learning methods within the same dataset conditions.

Separately, a threshold-based alert algorithm was developed to monitor breast milk temperature during transportation. This method triggers alerts only when three consecutive readings exceed a defined threshold, minimizing false positives. The algorithm also considers sensor activity to ensure alerts correspond to actual transportation events. Real-time monitoring is supported through periodic data acquisition and transmission via GPRS or CAT-M1 to a cloud platform.

In the context of chemometric analysis, one study used Partial Least Squares (PLS) regression to build calibration models based on Near-Infrared Spectroscopy (NIRS) data of breast milk. This was typically preceded by Principal Component Analysis (PCA) to identify and remove spectral outliers, improving model robustness. While traditional statistical and regression approaches remain relevant for calibration tasks, the inclusion of ensemble and learning-based methods reflects a broader interest in leveraging advanced data-driven tools for breast milk analysis and classification.

TABLE VI: Classification algorithms used for classifying breastmilk

References	Algorithm	Count
[12]	Random Forest (RF)	1
[12]	Support Vector Machine (SVM)	1
[12]	Adaboost	1
[12]	Back Propagation Neural Network (BP Neural Net)	1
[15]	Threshold-Based Alert Algorithm	1
[26]	Partial Least Squares (PLS) Regression	1
[26]	Principal Component Analysis (PCA)	1

What types and how many sensors are typically used in breastmilk classification? (RQ3)

In studies related to breast milk quality monitoring, various sensor configurations have been employed, ranging from basic environmental sensors to advanced spectroscopic systems. Reference [4] presents a system that utilizes two sensors: the DHT11 and DHT22 temperature and humidity sensors. These sensors are commonly used due to their affordability and reliability in capturing basic environmental data, and in some cases, the system includes an Arduino-based temperature and humidity sensor (model unspecified) for real-time monitoring in cold storage environments.

In a more advanced setup, Reference [12] describes the use of a single integrated sensor module, the T9602 temperature and humidity sensor, embedded within the Digital Matter Eagle Logger device. This system is designed specifically for monitoring donor human milk during transportation. It collects data at regular intervals and transmits it via cellular networks to a cloud-based storage platform. Although technically part of a multi-function device that also includes GPS and binary open/close status detection, the main sensing component for environmental monitoring is a single temperature-humidity sensor.

Meanwhile, Reference [15] discusses the application of a Near-Infrared Spectroscopy (NIRS) sensor, also categorized as a single-sensor system. This sensor is utilized for the real-time, reagent-free analysis of breast milk macronutrients, including fat, protein, and carbohydrate content. Operating in the wavelength range of 1596 to 2396 nm, this portable and low-cost spectroscopic sensor allows for accurate classification of breast milk quality without the need for sample pretreatment.

These configurations illustrate the range of sensor setups in current research from simple dual-sensor designs for environmental monitoring to single, high-precision sensors for compositional analysis. The variation in sensor counts and type reflects the diverse goals of each system, whether focused on storage safety, supply chain monitoring, or nutritional profiling.

TABLE VII: Type of Sensors Applied

References	Number Sensor Applied	Type of Sensor
[15]	3	DHT11, DHT22, Arduino-based Temp & Humidity
[23]	1	T9602 Temperature/Humidity
[26]	1	Near-Infrared Spectroscopy (NIRS)

Several limitations remain in the application of multi-array sensors and machine learning for breastmilk classification. These include high cost of sensor arrays, challenges in calibration and maintenance, limited sensor lifespan, and computational requirements for real-time deployment. In addition, accessibility and acceptance in low-resource settings may hinder large-scale adoption.

Research Question 5-7

To answer RQ 5, Figure 8 illustrates the distribution of research articles across three main fields: Healthcare & Nutrition, and Microbiology. The analysis reveals that the Healthcare field dominates, with a total of 3 articles, indicating a strong research focus on healthcare applications such as breast milk storage systems, quality control, and IoT-based health monitoring solutions. The Nutrition field follows, with 2 articles, highlighting interest in analyzing breast milk composition and nutritional value using technologies like Near-Infrared Spectroscopy (NIRS). Microbiology has the lowest representation, with only 1 article, suggesting limited but emerging research into microbiota classification and the microbial quality of breast milk. This distribution shows that while all three fields contribute to the broader study of breast milk and healthcare remain the most extensively explored area in current literature. Figure 8 is the summary of the result and discussion based on RQs. Details to answer RQs 5-7 are recorded in Table 9.

The analysis reveals that the Healthcare field dominates, with a total of 3 articles, indicating a strong research focus on healthcare applications such as breast milk storage systems, quality control, and IoT-based health monitoring solutions.

The Nutrition field follows, with 2 articles, highlighting interest in analyzing breast milk composition and nutritional value using technologies like Near-Infrared Spectroscopy (NIRS). Microbiology has the lowest representation, with only 1 article, suggesting limited but emerging research into microbiota classification and the microbial quality of breast milk. This distribution shows that while all three fields contribute to the broader study of breast milk and healthcare remain the most extensively explored area in current literature. The summary of comparison can be found at Table X

TABLE IX Summary of final collected paper for metric, field, limitation and future trend

References	Number Sensor Applied	Type of Sensor	Limitation	Future Trend
[1]	-Accuracy -Precision -Recall -F1 Score	Microbiology	Data Imbalance: The study faced challenges due to the unbalanced distribution of the dataset, which affected the performance of some models, particularly the Adaboost model.	Precision Medicine: The findings of this study could pave the way for advancements in precision medicine, where dietary recommendations and health interventions are tailored based on ethnic backgrounds and microbiota profiles.

			Model Performance Variability: The BP neural network did not achieve optimal results due to its higher requirements for data samples and stricter parameter control during training. SVM Limitations: The SVM model exhibited the worst performance, likely due to its unsuitability for multi-classification problems.	Enhanced Machine Learning Techniques: Future research may explore more sophisticated machine learning algorithms or hybrid models to improve prediction accuracy and handle data imbalances more effectively. Broader Applications: The methodology developed in this study could be applied to other areas of health and nutrition, potentially leading to a better understanding of how microbiota influences health across different populations.
[2]	-Performance Indicators -Conceptual Model Validity	Healthcare	Lack of Literature: The authors note a significant gap in the literature regarding the application of CM frameworks to real-world case studies, which hinders the practical implementation of CM research. Uncertainty in Parameters: The paper discusses the challenges posed by the uncertainty of supply and demand parameters in real-world scenarios, which complicates inventory management for perishable goods like human milk.	Strategic Planning: The study aims to develop a model that supports strategic future planning and preparedness for new policy implementations in human milk banking. This indicates a trend towards using simulation models for long-term decision-making rather than immediate operational needs. Enhanced Communication Tools: The authors suggest the potential for using 2D animations in simulation software to communicate findings to a broader audience, indicating a trend towards more visual and accessible representations of complex models.
[4]	-Real-time monitoring	Healthcare and Nutrition	Existing Systems: Many milk banks	Broader Applications: The technology

			currently rely on outdated methods for monitoring storage conditions, which may not be as effective as the proposed IoT solution. This indicates a gap in technology adoption that could hinder the implementation of the new system. Scalability Challenges: While the proposed system has potential applications in other cold storage facilities, the actual scalability may face challenges due to varying infrastructure and resource availability in different regions.	developed could be adapted for use in other sectors, such as food banks and pharmacies, indicating a trend towards more widespread use of IoT solutions in various fields. Enhanced Monitoring Systems: Future developments may focus on improving the accuracy and reliability of monitoring systems, potentially incorporating advanced sensors and data analytics to further enhance operational efficiency. Integration with Other Health Initiatives: There is potential for further integration of IoT solutions with other health initiatives, which could lead to improved health outcomes for mothers and infants.
[12]	-Temperature and Humidity Monitoring -Quality Control Alerts	Healthcare	Sensor Battery Life: One significant limitation noted in the study is the need to frequently change the batteries of the sensors. This process can be time-consuming, especially when managing a large number of sensors, which may affect the continuous use of the technology. Sensor Retrieval Issues: The study mentions challenges in retrieving sensors from the premises of the human milk bank, which can hinder the	Expansion to Other Clinical Resources: The paper suggests that the use of IoT sensors in monitoring DHM can be extended to other critical supply chains, such as those for blood and human organs. This could lead to improved quality control and reduced waste across various medical fields. Sustainability Initiatives: As the world moves towards sustainability, the integration of IoT technology in supply chains is expected to play

			effectiveness of the monitoring system.	a vital role in reducing costs and improving energy efficiency, thereby supporting sustainable practices in healthcare.
[15]	-Correlation Coefficient (r^2) -Standard Error of Cross-Validation (SECV) -Ratio of Performance to Deviation (RPD)	Nutrition	Water Interference: Strong water absorption in the NIR region increases background noise, making quantitative analysis more difficult. Device Sensitivity: Low-cost NIRS devices often have narrow wavelength ranges and lower sensitivity, which can reduce accuracy. Spectral Complexity: Overlapping molecular vibrations in NIR spectra make interpretation challenging, requiring complex multivariate analysis.	Data Analysis Enhancement: Improvements in multivariate analysis techniques will enhance the interpretation of complex NIR data for better macronutrient prediction. Device Innovation: Development of advanced portable NIRS devices with wider wavelength ranges and higher sensitivity to increase accuracy and usability. Aquaphotomics Focus: Growing interest in aquaphotomics to understand water's role in NIR spectra.

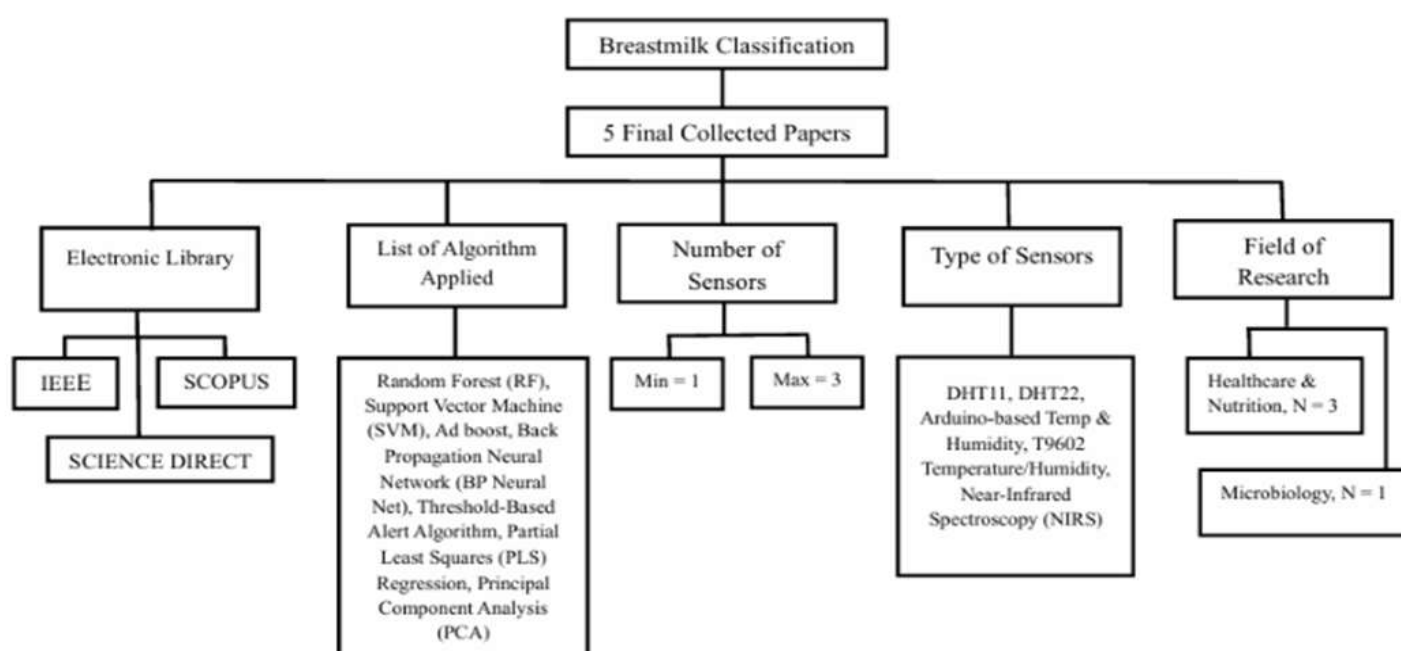


Fig. 6 Summarization of results and discussion based RQs

TABLE X Comparative Summary Table

Approach	Key Methods	Strengths	Limitations	Typical Use Case
Traditional	Spectroscopy, chemical assays	Well-established, accurate	Time-consuming, costly, not real-time	Lab-based analysis
Machine Learning	RF, SVM, AdaBoost, BPNN	Balance of accuracy & efficiency, works with smaller datasets	Still requires feature engineering, dataset imbalance	Sensor-based classification
Deep Learning	CNN, LSTM (not widely applied yet in breastmilk studies)	High potential accuracy, automatic feature extraction	Requires large datasets, high computational cost	Future large-scale IoT integration

Limitation

Despite extending our search to four major databases (Scopus, IEEE Xplore, ScienceDirect, and Google Scholar) and considering grey literature, only five studies met the stringent inclusion and quality assessment criteria set in this review. This relatively small number of eligible studies underscores that research on breastmilk quality classification using multi-array sensor technology and machine learning remains at an early stage of development. While the limited dataset may restrict the generalizability of the findings, it also highlights the novelty and originality of this review, since it systematically consolidates evidence from a domain that has received little scholarly attention.

In fact, the scarcity of directly related studies reinforces the importance of continued investigation in this area, and it signals a substantial research gap that future scholars and practitioners can address. Furthermore, the small pool of studies may reflect broader challenges, such as the cost of conducting empirical validation, technical barriers in deploying multi-sensor platforms, and limited accessibility of large-scale breastmilk datasets. These contextual limitations should be acknowledged, as they provide critical insight into why this research field remains underexplored, while also emphasizing the significance of the current review in laying the groundwork for future empirical and applied research.

Future Direction

Future studies should focus on integrating sensor systems with IoT and cloud platforms to enable remote, real-time monitoring, which would significantly enhance both scalability and accessibility of breastmilk quality assessment. Such integration would allow continuous data collection, storage, and analysis across multiple locations, thereby reducing reliance on conventional laboratory testing.

At the same time, addressing data privacy and security will be critical for safeguarding sensitive maternal and infant health information, especially when cloud-based or mobile applications are employed. Ensuring compliance with international standards such as the General Data Protection Regulation (GDPR) in the European Union or the Health Insurance Portability and Accountability Act (HIPAA) in the United States will strengthen user trust and facilitate adoption in both clinical and personal settings.

Moreover, the development of low-cost, portable, and user-friendly devices is essential to ensure that this technology is not restricted to advanced healthcare facilities but can also be deployed in resource-limited environments and households. User-centered design and inclusive testing should be prioritized so that the devices remain practical and acceptable for everyday use by mothers and healthcare providers. Beyond technical innovation, interdisciplinary collaboration involving healthcare professionals, engineers, policymakers, and social scientists will be required to address challenges such as sensor calibration, device maintenance, and long-term sustainability. By tackling these aspects, future research can create more comprehensive, equitable, and impactful solutions for improving maternal and infant healthcare worldwide.

CONCLUSION

This systematic literature review has highlighted the emerging role of machine learning algorithms combined with multi-array sensor technologies in advancing breastmilk quality classification. The findings reveal that while traditional laboratory-based methods remain reliable, they are often limited by cost, time, and practicality for real-time applications. Machine learning techniques such as Random Forest and Support Vector Machines demonstrate strong potential due to their balance of accuracy and efficiency, whereas deep learning approaches, though promising, require larger datasets and higher computational resources.

The review also shows that feature extraction, sensor integration, and algorithm selection are critical factors influencing classification outcomes. Current research, however, remains scarce, with only a handful of studies directly addressing this domain. This indicates both a significant research gap and a promising opportunity for future exploration.

Moving forward, efforts should focus on developing robust, scalable, and low-cost systems that integrate intelligent algorithms with sensor-based platforms for real-time monitoring of breastmilk quality. Such innovations could transform maternal and infant healthcare by enabling safer, more accurate, and more accessible quality assessment methods. Ultimately, this work lays a foundation for future research and practical applications, contributing to the advancement of smart healthcare solutions for mothers and infants worldwide.

ACKNOWLEDGMENT

This study was supported by research from University Technical Malaysia Melaka (UTeM). The authors would like to express their sincere gratitude to colleagues at the Faculty Technology Maklumat & Komunikasi, UTeM, for their valuable technical input and constructive feedback during the development of this work. The authorship of this article reflects equal contribution from all authors involved in the study. Special thanks are extended to individuals who assisted in formatting and proofreading the manuscript.

REFERENCES

1. R. A. Lawrence, "Storage of human milk and the influence of procedures on immunological components of human milk," *Acta Paediatrica, International Journal of Paediatrics, Supplement*, vol. 88, no. 430, pp. 14–18, 1999, doi: 10.1111/j.1651-2227.1999.tb01295.x.
2. C. G. Victora et al., "Breastfeeding in the 21st century: Epidemiology, mechanisms, and lifelong effect," *The Lancet*, vol. 387, no. 10017, pp. 475–490, 2016, doi: 10.1016/S0140-6736(15)01024-7.
3. M. Naquiuddin, A. Ibrahim, M. Sharfi Najib, and S. M. Daud, "MEKATRONIKA JOURNAL OF MECHATRONICS AND INTELLIGENT MANUFACTURING Case Modelling Odour Profiles and Temperature Intensity of Water: A Comparative Analysis using Case-Based Reasoning and K-Nearest Neighbours," vol. 6, no. 2, 2024, doi: 10.15282/mekatronika.v6i2.10729.
4. H. Yang, W. Jiao, L. Zouyi, H. Diao, and S. Xia, *Artificial intelligence in the food industry: innovations and applications*, vol. 5, no. 1. Springer International Publishing, 2025. doi: 10.1007/s44163-025-00296-8.
5. S. Narendra Kumar, A. Patil, D. R. Reddy, R. U. Kushal, and P. Basavaraj, "Predictive Analytics for Milk Quality Using Random Forest (RF) Algorithm," 8th IEEE International Conference on Computational System and Information Technology for Sustainable Solutions, CSITSS 2024, no. November, pp. 1–5, 2024, doi: 10.1109/CSITSS64042.2024.10816917.
6. R. Serreau, Y. Terbeche, and V. Rigourd, "Pollutants in Breast Milk: A Scoping Review of the Most Recent Data in 2024," *Healthcare (Switzerland)*, vol. 12, no. 6, 2024, doi: 10.3390/healthcare12060680.
7. R. Li, N. Shenker, J. Gray, J. Megaw, G. Weaver, and S. J. Cameron, "Microbiological analysis of donor human milk over seven years from the Hearts Milk Bank (United Kingdom)," *Food Microbiology*, vol. 126, no. August 2024, p. 104661, 2025, doi: 10.1016/j.fm.2024.104661.
8. S. Janakiraman, R. Sha, and N. K. Mani, "Recent advancements in Point-of-Care Detection of Contaminants and Biomarkers in Human Breast Milk: A comprehensive review," *Sensors and*

- Actuators Reports, vol. 9, no. September 2024, p. 100280, 2025, doi: 10.1016/j.snr.2024.100280.
9. M. M. Brambilla et al., “Systematic review on Endocrine Disrupting Chemicals in breastmilk and neuro-behavioral development: Insight into the early ages of life,” *Neuroscience and Biobehavioral Reviews*, vol. 169, no. September 2024, p. 106028, 2025, doi: 10.1016/j.neubiorev.2025.106028.
10. Y. Wang, K. Zhang, S. Shi, Q. Wang, and S. Liu, “Portable Protein and Fat Detector in Milk Based on Multi-Spectral Sensor and Machine Learning,” *Applied Sciences (Switzerland)*, vol. 13, no. 22, 2023, doi: 10.3390/app132212320.
11. S. Agudelo-Pérez, D. Botero-Rosas, L. Rodríguez-Alvarado, J. Espitia-Angel, and L. Raigoso-Díaz, “Artificial intelligence applied to the study of human milk and breastfeeding: a scoping review,” *International Breastfeeding Journal*, vol. 19, no. 1, pp. 1–15, 2024, doi: 10.1186/s13006-024-00686-1.
12. X. Zhou, W. Yan, Y. Ren, Q. Zhao, and Y. Zheng, “A Machine Learning-Based Model for Predicting Breast Milk Flora Ethnicity,” *Proceedings of 2023 7th Asian Conference on Artificial Intelligence Technology, ACAIT 2023*, pp. 21–27, 2023, doi: 10.1109/ACAIT60137.2023.10528608.
13. M. Staff, N. Mustafee, and N. Shenker, “Conceptual Modeling for Perishable Inventory: A Case Study in Human Milk Banking,” *Proceedings - Winter Simulation Conference*, pp. 1208–1219, 2023, doi: 10.1109/WSC60868.2023.10407264.
14. R. Tharun, G. S. Pavithra, V. Sreekanth, N. S. Sundar, and T. G. Harshitha, “Design of Sustainable Automated Milking and Milk Quality Testing Machine,” *2024 5th International Conference for Emerging Technology, INCET 2024*, pp. 1–6, 2024, doi: 10.1109/INCET61516.2024.10593149.
15. K. Phathela and A. J. Henney, “Internet of Things Application in South African Breast Milk Banks,” *2024 Conference on Information Communication Technology and Society, ICTAS 2024 - Proceedings*, pp. 139–143, 2024, doi: 10.1109/ICTAS59620.2024.10507145.
16. A. Deshpande, S. Deshpande, and S. Dhande, “NIR Spectroscopy Based Milk Classification and Purity Prediction,” *2021 IEEE Pune Section International Conference, PuneCon 2021*, pp. 1–5, 2021, doi: 10.1109/PuneCon52575.2021.9686473.
17. J. L. C. Paulino et al., “Milk Bank PH: A Website for Donating and Reserving Breastmilk for Infants,” *Proceedings - 2021 1st International Conference in Information and Computing Research, iCORE 2021*, pp. 198–204, 2021, doi: 10.1109/iCORE54267.2021.00054.
18. P. Suaprae, T. Janchidfah, and B. Bhatt, “Comparison of Milk Quality Grades Using the Logistic Model Tree Algorithm,” *International Conference on Cybernetics and Innovations, ICCI 2025*, pp. 1–7, 2025, doi: 10.1109/ICCI64209.2025.10987482.
19. W. Wang et al., “Characterization and classification of odorous raw milk: Volatile profiles and algorithm model perspectives,” *Journal of Food Composition and Analysis*, vol. 138, no. November 2024, p. 107030, 2025, doi: 10.1016/j.jfca.2024.107030.
20. M. Bovo, M. Agrusti, S. Benni, D. Torreggiani, and P. Tassinari, “Random forest modelling of milk yield of dairy cows under heat stress conditions,” *Animals*, vol. 11, no. 5, 2021, doi: 10.3390/ani11051305.
21. E. M. Morse-McNabb, M. F. Hasan, and S. Karunaratne, “A Multi-Variable Sentinel-2 Random Forest Machine Learning Model Approach to Predicting Perennial Ryegrass Biomass in Commercial Dairy Farms in Southeast Australia,” *Remote Sensing*, vol. 15, no. 11, 2023, doi: 10.3390/rs15112915.
22. Y. Zhang, L. Zhang, Y. Ma, J. Guan, Z. Liu, and J. Liu, “Research on dairy products detection based on machine learning algorithm,” *MATEC Web of Conferences*, vol. 355, p. 03008, 2022, doi: 10.1051/mateconf/202235503008.
23. U. Ramanathan, K. Pelc, T. P. da Costa, R. Ramanathan, and N. Shenker, “A Case Study of Human Milk Banking with Focus on the Role of IoT Sensor Technology,” *Sustainability (Switzerland)*, vol. 15, no. 1, 2023, doi: 10.3390/su15010243.
24. M. Muelbert, F. H. Bloomfield, S. Pundir, J. E. Harding, and C. Pook, “Olfactory Cues in Infant Feeds: Volatile Profiles of Different Milks Fed to Preterm Infants,” *Frontiers in Nutrition*, vol. 7, no. January, 2021, doi: 10.3389/fnut.2020.603090.
25. Nayana MS, Nekkanti Deepak, Nisha M, Shravani MS, and Dr. Manjunath HR, “A Review on a Milk Quality Detection and Analysis,” *International Journal of Advanced Research in Science, Communication and Technology*, vol. 3, no. 1, pp. 76–79, 2023, doi: 10.48175/ijarsct-7838.

26. C. Melendreras et al., “Near-Infrared Sensors for Onsite and Noninvasive Quantification of Macronutrients in Breast Milk,” *Sensors*, vol. 22, no. 4, pp. 1–12, 2022, doi: 10.3390/s22041311.
27. N. J. Kannampilly, K. Thangavel, D. Peter, and L. Rose, “Milk spoilage detection by impedance measurement,” *International Journal of Current Research and Review*, vol. 13, no. 5, pp. 183–187, 2021, doi: 10.31782/IJCRR.2021.13534.