2023 11th International Conference on Emerging Trends in Engineering & Technology - Signal and Information Processing (ICETET - SIP)

Deep Learning Based Approach for Milk Quality Prediction

Saumya Kumari
School of Computer Engineering
KIIT Deemed to be University
Bhubaneswar 751024, Odisha
Saumyastaofficia
l@gmail.com

Mahendra Kumar Gourisaria
School of Computer Engineering
KIIT Deemed to be University
Bhubaneswar 751024, Odisha
mkgourisaria2010@gmail.com

Himansu Das
School of Computer Engineering
KIIT Deemed to be University
Bhubaneswar 751024, Odisha
himanshufcs@kiit.ac.in

Debajyoty Banik
School of Computer Engineering
KIIT Deemed to be University
Bhubaneswar 751024, Odisha
debajyoty.banik@gmail.com

Abstract. **This study aimed to find out what factors have the biggest impact on milk quality. We looked at 9 different variables including pH, temperature, taste, odor, fat, turbidity, color, and grade to see which ones were the most important. We used the method called Principal Component Analysis (PCA) and found that temperature and color were the top factors affecting milk quality, with over 95% contribution to PCA-1 and PCA-2. We also divided the milk samples into three grades - low, medium, and high - and used a machine learning algorithm called Artificial Neural Network (ANN) to classify the milk samples. We found that the ANN model was able to classify the milk samples with high accuracy (0.9988) and was more accurate and stable compared to other methods. Other advanced deep learning algorithms could have been used to improve the results further.**

Keywords: quality milk, statistical technology approach, dairy industry, support vector

I. INTRODUCTION

The quality of milk consumed and sold is often compromised by various contaminants such as detergents and skim milk powders, but impure water is the most prevalent. The Food Safety and Standards Authority of India (FSSAI) conducted a National Survey on Milk Adulteration in 2011, which revealed that water was the most common adulterant in Indian milk, followed by detergent. To address this issue and safeguard the health of consumers, the Indian government introduced the Prevention of Food Adulteration Act (PFA Act) in 1954, which came into effect on June 1, 1955 [1]. The production, sale, and distribution of contaminated and poisonous foods are prohibited by law. Despite legislative efforts, fortification persists due to a shortage of well-trained personnel and inadequate laboratories. To evaluate the quality of food products, various studies have been conducted using data from reputable sources and employing machine learning techniques. This shows that it is possible to use the neural network technology to ensure that milk produced is of high quality based on numerical data. The potential of new technologies will be realized if new approaches are identified to manage milk quality using neural network technologies [2], thereby saving resources and preventing health hazards from occurring. Using numeric data, machine learning methods such as neural networks and decision trees can be utilized to develop models that embody the properties of both high-grade and adulterated milk samples. Through precise detection, the expenses associated with treatment can be diminished, the spread of illness caused by substandard milk can be curtailed, and the quality of milk can be sustained. Efficient detection not only reduces costs but also ensures the well-being of cattle by detecting diseases or infections in advance. The use of a computational model helps farmers implement appropriate treatment regimens in a timely manner; thereby preventing the spread of illness to healthy cattle [3].The objective of this research is to improve the precision of milk predictive analysis. To achieve this, Machine Learning (ML) techniques are employed instead of traditional methods such as cluster analysis and discriminate analysis. The data is partitioned into training and testing sets, which are used to train the ML algorithms. This enables the computer to refine its predictions by learning from previous results.

In this paper, we present a comprehensive evaluation of various machine learning algorithms applied to classify the quality of milk. The methods examined include regression techniques and classifiers such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Random Forest (RF). We also introduce the use of Artificial Neural Networks (ANN) and compare its performance against the other algorithms. The results indicate that the ANN model achieved a superior balance of computational efficiency, accuracy, and stability, making it the most appropriate for milk quality classification.

2023 11th International Conference on Emerging Trends in Engineering & Technology - Signal and Information Processing (ICETET - SIP) | 979-8-3503-4842-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICETET-SIP58143.2023.10151626

II. RELATED WORK

Classifying milk samples is crucial for maintaining the quality of the dairy product delivered from farms. Due to the presence of various chemical substances and physical factors that impact the quality of milk, it is difficult to establish a consistent classification method. Numerous studies have been conducted to find an effective solution to this classification problem.

In a recent study by Olcay Polat (2021) [4], the use of an information fusion framework was explored as a means of classifying raw milk samples. The aim was to develop a suitable approach to categorize raw milk into multiple classes based on important criteria. The study results showed that pH, sH and somatic cell count was significant factors in determining the quality of raw milk. In the field of dairy processing, Multi-criteria Decision-Making (MCDM) and Analytic Hierarchy Process (AHP) techniques were implemented, but their computation complexity was high. To address this issue, Pegah Sadeghi Vasafi (2021) [5] utilized KNN and SVM classification techniques to identify anomalies in the milk processing process. The fat content and temperature were considered as features and the accuracy achieved was 81.4% for SVM and 84.8% for KNN. Despite the success of these two techniques, a higher accuracy could have been achieved by incorporating additional classifiers. W Habsari (2021) [6] aimed to develop a smart grading system for milk quality classification in the dairy industry using ANN and K-means models. The factors affecting milk quality, such as pH, temperature, odor, turbidity, color, fat and taste values, were considered in the study. The accuracy obtained was 98.74%, with color and temperature being the attributes used for grouping. Although the clusters produced by the K-mean algorithm were limited, no other models were applied to cross-validate the results, except for the Artificial Neural Network model. The application of Artificial Neural Network (ANN) for predicting the yield of milk in dairy farms has been a topic of interest for researchers. In 2013, Shailesh Chaturvedi [7] employed ANN model to predict the milk amount and constructed two hidden layers of the network. The model's performance was compared to actual experimental data and showed that ANN could be used to determine the future performance of dairy cattle based on early traits expressed. The network consisted of four input layers, two hidden layers, and one output layer. 60% of the data was used for training and 20% each for testing and verification. However, the error could have been reduced and the correlation coefficient maximized closer to 1.

Earlier, in 2002, A.P. Kominakis [8] studied the use of ANN for predicting the yield of milk and found that network specialization improved the prediction ability, resulting in a high correlation between observed and predicted values (0.87 to 0.97). However, the standard deviation of the data was misjudged, leading to an unexpectedly high standard deviation (s=91.55).

Machine learning and deep learning have made a significant impact in various fields, including agriculture. Despite numerous attempts to classify milk quality and predict factors associated with it, there is a lack of proper evaluation in most cases. The delivery of low-grade milk can lead to illnesses and nutritional deficiencies in the population.

III. METHODS APPLIED

This section provides detail about the preparation of the data set and the technologies used to get the desired result. The data set had no null or duplicate values present in it. Though in order to yield better results both regression as well as the neural model made and then the accuracy was compared. Principal Component Analysis (PCA) which is a useful tool for determining which quality parameters influenced the overall quality, and other machine learning algorithms were also used to construct the classification models, like Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbour (KNN) and Random Forest (RF). Under the neural network model, the artificial neural Network (ANN) model was built.

A. Qualitative analysis of milk by principal component analysis

A principal component analysis (PCA) is a method of producing linearly uncorrelated variables from several correlated variables. A data set is most accurately described by its first principal component in this conversion. By using only, the first few principal components, PCA reduces the dimension of data set so that it can be transformed to a lower-dimensional data set. In essence, PCA finds a new base of orthogonality.

Table 1. Tabular analysis of composition of all features to PCA-1 and PCA-2

Parameters	PCA-1	PCA-2
pH	0.0345	0.0580
Temperature	0.9993	-0.0065
Taste	-0.00542	0.009625
Odour	-0.0024	0.0043
Fat	0.001	-0.012
Turbidity	0.00916	-0.01621
Color	-0.0047	-0.9980

KDE (Kernel Distribution Estimation) plots of features were unable to determine the important features to reduce dimensionality. So, we applied the Principal Component Analysis, which showed that the temperature and colour have over 95% composition to PCA-1 and PCA-2 as shown in Table 1. Hence, in order to determine the milk's quality the milk factory should prioritize the parameters temperature and colour.

B. *Artificial neural network based milk quality detection*

In information processing, an Artificial Neural Network (ANN) mimics brain function. Through a learning process, an ANN is configured for specific applications, such as pattern recognition or data classification. Research has proven to be effective in a wide range of disciplines, from medicine to engineering, by using artificial neural networks-based approach in complex analysis [9-12]. A neural network consists of hundreds of nodes, or artificial neurons, connected by coefficients (weights), forming the neural structure. Neural computations are powered by networks of neurons. There is one output and one weighted input for each PE. Transfer functions, learning rules, and architecture determine the behaviour of a neural network. In that sense, a neural network is a parameter system, since the weights are adjustable. It is the weighted sum of the inputs that determines whether the neuron is activated. The activation signal is passed through a transfer function to produce a single output of the neuron. The model works in the following way:

An input unit is passed with some weights attached to it to the hidden layer in the first step. The number of hidden layers is not limited. X1, X2, X3 ... Xn are the inputs. Neurons are present in each hidden layer. Neurons are connected to all inputs. As soon as the inputs are passed on to the hidden layer, all computation occurs there.
The Calculations in hidden layers are performed in two steps as follows:

To begin with, each input is multiplied by its weight. Variables are weighted according to their gradients or Coefficients. This indicates how strong a particular input is. Adding the bias variable comes after assigning weights. Adding bias to a model will ensure that it fits as accurately as possible as shown in Equation 1 below.

$$Z1 = \sum_{i=1}^{n} W_i * ln_i + b \qquad (1)$$

Where W_i (I=1, 2, 3…, n) are the weights assigned to the inputs ln_i (I=1, 2, 3…, n) and b is the bias.

A linear equation denoted by Z1 here is then activated by applying an activation function. Before sending input to the next layer of neurons, the activation function transforms it nonlinearly. Activation functions instil nonlinearity in models. There are several activation functions such as Sigmoid, ReLu, LeakyReLU, PReLU, ELU, Tanh, and SoftMax.

Every hidden layer undergoes the above process. After passing through each hidden layer, which is our output layer, we arrive at the final output. We call this process forwarding propagation. The error is calculated based on the predicted output and the actual output, which is the difference between the two. We use Back Propagation when the error is large in order to minimize it.

C. *Support vector machine-based milk quality identifier*

SVM can also be used for regression analysis. The objective function is to find the hyper plane that divides the classes correctly. In case of 2-D data the decision boundary becomes a straight line rather than the hyper plane. The best of hyper plane is decided on the basis of its distance being maximum from the present classes and this serves as the prime objective of SVM.The equation of hyper plane is mentioned below as Equation 2:

$$\vec{a}.\vec{x} + b = 0 \qquad (2)$$

where, $\vec{a}$ is a vector normal to the hyper plane, $\vec{x}$ is a vector to any random point assumed which can be above, below or on the hyper plane and b is an offset. In order to determine the location of the random point assumed, we have a decision rule which says that if the above-stated equation is greater than or equal to zero, the point is positive and lies below the hyper plane. Whereas, if the equation equates to be less than zero then the point lies above the hyper plane being negative. The stated fact can be summed up and written in the mathematical format as shown in Equation 3.

$$f = \begin{cases} (+1) \text{ if } \vec{a}.\vec{x} + b \geq 0 \\ (-1) \; if \; \vec{a}.\vec{x} + b < 0 \end{cases} \qquad (3)$$

For optimization of the hyper plane, we need to decide on a constraint and the constraint should be such that while calculating the distance of the closest observations to the hyper plane, it should be maximum and not any other observed point should lie beyond this closest observation's margin line. If we consider variable 'd' to be as that distance, then in order to maximize d and obtain best results the constraint should hold true as stated below in Equation 4.

$$f_i(\vec{a}.\vec{x} + b) \geq 1 \qquad (4)$$

The above-stated constraint irrespective of the point being positive or negative needs to be true always if we want to maximize d. Now, we consider two support vectors, each being from both positive and negative classes. Let X_1 and X_2 be those two support vectors such that the distance between them should be shortest and the projection of difference of these two vectors on a vector perpendicular to the hyper plane which was taken as a vector $\vec{a}$ will help to obtain optimization function as shown in Equation 5.

$$\text{argmax}(a^*, b^*) \frac{2}{||a||} \text{ for } f_i(\vec{a}.\vec{x} + b) \geq 1 \quad (5)$$

Although at times when the data are not easily linearly separable the sated optimization functions need modification. We need to add a soft margin which leads to the addition of a few more terms in the equation. Firstly, we invert the max function in the min format and then add the term zeta ('z') to the equation multiplied with hyper parameter 'c'. The zeta is nothing but an additional value that equals zero when all the points are correctly classified but equals the distance of the

incorrect point from its correct hyper plane in the other case. The value of hyper parameter 'c' holds an inverse relationship with classification error and hence in order to decrease the classification error, the value of c should be increased. Therefore, the modified optimization function for the hyper plane will be as shown in Equation 6.

$$\text{argmin } (a^*, b^*) \frac{||a||}{2} + c \sum_{i=1}^{n} z_i \text{ for } f_i(\vec{a}.\vec{x} + b) \geq 1 \quad (6)$$

D. Quality assessment of milk using random forest

Random forest methods, which are bagging-based, facilitate ensemble learning. The model has a strong anti-noise ability and can avoid over fitting when solving classification and regression problems. A new subset of training points is created by selecting m sample points from the training sample set S; a classification decision tree or regression model is then constructed for each training subset by randomly selecting n features from all features as split nodes; the average output of each decision tree is the category with the most votes, or the average output (regression).

As we know that the method can be applied to solving regression and classification problem, and the equation will be derived differently for both cases. In case of regression, the equation for Mean Squared Error (MSE) will be as shown in Equation 7.

$$X = \frac{1}{m} \sum_{i=1}^{m} (f_i - y_i)^2 \quad (7)$$

Where, X is the mean squared error; m is the number of sample points taken from the training sample set S, f_i is the value given by each model and y_i is the actual value for each tree i. The stated equation helps to estimate the distance of each tree from the predicted actual value and hence in picking out the suitable branch for our decision tree.

In the case of the classification problem, the equation for Gini will be as demonstrated in Equation 8.

$$X = 1 - \sum_{i=1}^{n} t_i \quad (8)$$

Here, X is the Gini to calculate with n representing the classes present and t_i the relative frequency of the present class from the data set being observed. In place of Gini, we can also use entropy method where we take X as the entropy to be calculated which helps to determine the way the tree branch from the decision tree. The equation gets modified as represented in Equation 9.

$$X = \sum_{i=1}^{m} - l_i * \log_2 t_i \quad (9)$$

E. K-nearest neighbour-based milk quality identifier

The k-nearest neighbour's (KNN) is another efficient machine learning algorithm used for data classification. It helps in estimating the probability of whether a random data point will belong to one or the other group on the basis of checking the belongingness of the data points nearest to it. It can be used for solving both classification and regression problems but is more helpful in the case of a prior one [13].

While choosing the K value, we should try to pick up an odd value for it. There are several distance methods available with the help of which nearest or neighbourhood points can be calculated.

Using Euclidean Distance, the equation formed will be as stated in Equation 10.

$$d = \sqrt{\sum_{i=1}^{n} (y1_i - y2_i)^2} \quad (10)$$

Where d is the Euclidean distance and (y1-y2) gives the shortest distance between two points. The method is used when data of high dimension is not to be considered.

Although in the case of high dimensional data and especially when it is a grid type of data, we use Manhattan Distance which is given as stated in Equation 11.

$$d = \sum_{i=1}^{n} |y1_i - y2_i| \quad (11)$$

Here, the d represents the Manhattan distance and (y1-y2) gives the shortest distance between two points.

F. *Quality assessment of milk using logistic regression*

In machine learning, logistic regression is a supervised algorithm for identifying patterns in data. Numerous researchers in the past have used a logistic regressions-based approach to predict the target variable's probability and to confound the data [14-18]. Logistic Regression helps in trainability, easy implementation and ease of interpretation are the advantages of logistic regression and no assumptions are made about the distribution of classes in feature space but in case of the number of features are more than the number of observations, it might lead to over-fitting and hence is not recommendable in such cases. Further, it also builds up linear boundaries.

In order to solve the classification problem, it must find the minimum loss function value [19]. Hence, in order to compare the machine learning algorithm's performance this technique was used.

IV. DATA-SET DESCRIPTION

The data set used to predict the quality and grade of milk was taken from Kaggle. The dimension of the data set used is 1059 rows x 9 columns as described in Table 2. As claimed about the data set, it was manually collected from

observations to help us in building machine learning models to predict milk quality. It consists of seven independent variables - pH, Temperature, Taste, Odour, Fat, Turbidity, and Colour. Generally, these parameters determine the milk's quality or grade. The column pH contains information about the pH (potential of hydrogen) value of the milk produced and its value ranged from 3 to 9.5. This value plays a key role in detecting the impurities and signs of infection in the milk produced. Then comes the 'Temperature' column which contains information about the temperature with values ranging from 34 degrees Celsius to 90 degrees Celsius, at which the milk produced is maintained so that the quality of milk can be preserved by keeping a check on it. It is scientifically proven that there are high chances of the growth of bacteria when the temperature goes below 45 degrees Fahrenheit. The next columns, 'Taste', 'Odour', 'Fat' and 'Turbidity' contains information about the taste, smell, fat content and effective heating of milk till its sterilization point respectively. Although these columns store values in binary format which mean 1 or 0 rather than numerical values. This indicates there is a choice between either good or bad. Here we are considering 1 as good and 0 as a bad quality for the milk produced. Later comes the 'Colour' column which is an indicator of physio-chemical changes taking place in the milk produced. The values in the colour column are ranging from 240 to 255. Fat globules are suspended in a temporary emulsion in milk, which also contains colloidal albumin, salts, and sugar. Depending on the reason, its colour may go from yellowish white to nearly white to bluish-white tinted. At last, the 'Grade' column has three classes which are high, medium and low determining the quality of milk produced. There is no data pre-processing or cleansing performed due to the absence of any null values and hence other techniques like Principal component analysis (PCA) are used to reduce the dimensions of model inputs for the development of classification models (or milk's quality predictive analysis model).

Table 2. Details about the data set used having 1059 rows x 9 columns

Parameters	Description of the milk samples provided by them:
pH	Potential of hydrogen value in the milk Milk samples of the data set.
Temperature	Value at which the milk samples are stored.
Taste	Binary format value where 1 is for good taste and 0 for bad taste.
Odour	Binary format value where 1 is for good odour and 0 for bad odour.
Fat	Binary format value where 1 is for high-fat content and 0 for low-fat content.
Turbidity	Binary format value where 1 is for high turbidity and 0 for low turbidity.
Colour	Numerical values for the colour of the milk samples. White, yellowish white and slight bluish-white indicates good milk samples.
Grade	Three classes which are high, medium and low determine the quality of milk produced.

V. RESULTS AND ANALYSIS

We executed all the machine learning algorithms as mentioned in section 3 above. The entire dimension of the data set which is 1059 rows x 7 columns was taken for training, testing and evaluation. We performed a data visualization exercise on the target variable, 'Grade', where each feature was thoroughly explored and analyzed. The results indicated that milk samples with temperatures exceeding 45 degrees, pH values between 6 and 7, and low-fat content were of low quality. Then we used regression tools like logistic regression and SVM, RF and KNN. The different metrics evaluation sections like Precision, F1-score, Recall, Kappa and Jaccard score are mentioned in Table 3. In terms of classifier accuracy scores, both Random Forest (RF) and K-Nearest Neighbor (KNN) achieved a score of 0.9968, while Logistic Regression (LR) and Support Vector Machine (SVM) showed a score of 0.8490 and 0.9528 respectively. The precision score, which measures the number of positive class predictions that are actually positive, was used to determine the accuracy of the minority class in an imbalanced classification. The precision score increased with an increase in the minority class.

The recall, which refers to the number of positive class predictions achieved, and the F1-score, which strikes a balance between precision and recall, also increased in the order of LR, SVM, RF, and KNN. The F1-score, which combines both precision and recall, is a useful metric to achieve successful classification. The F1 scores obtained ranged between 0.8 to 0.9, indicating a successful classification of almost all observations into the correct class by the models. KNN and RF classifiers showed particularly better results.

Table 3. Metrics Evaluation of models used

Model	LR	SVM	RF	KNN
Accuracy	0.8490	0.9528	0.9968	0.9968
F1 Score	0.8416	0.94803	0.9963	0.9963
Precision	0.8397	0.9452	0.9956	0.9956
Recall Score	0.8487	0.9515	0.9971	0.9971
Jaccard	0.7293	0.9036	0.9927	0.9927
Kappa Score	0.7710	0.9279	0.9951	0.9951

Secondly, in order to build neural network model we proceeded with encoding the categorical data. Since our 'Grade' column had string values, the respective three classes which were high, low and medium were assigned the numerical values 0, 1 and 2 respectively. Then the splitting of the data set was performed with the test value 0.2 and splitting ratio 6:1. Later we proceeded with feature scaling where the Standard scalar was used and finally the ANN model was built. We had {20, 15, 10, and 3} layers and the epoch value were set to 50. The first 3 layers had an

activation function as ReLu, and the last output layer had SoftMax activation. For Artificial Neural Network, the Batch size was set to 10 and Adaptive moment estimation (Adam) was used as an optimizer.
Furthermore, after setting the epoch value to 50, the artificial neural network model achieved an accuracy score of 0.9988, demonstrating the reliability and efficiency of neural network models in dealing with classification problems compared to regression methods.
Finally, the Principal Component Analysis revealed that temperature and color were the most important factors for determining the quality of milk produced in the milk factory.

VI. CONCLUSION AND FUTURE WORK

We made use of a machine learning algorithm in this paper to make a predictive analysis of the quality of milk with the help of some general factors which affect its quality. It becomes important to maintain the quality of milk so that the related dairy products manufactured can also be of the best quality as well as the outburst of any kind of disease or allergy can be prevented. Since the demand for milk consumption is growing day by day with an increase in the population, it becomes a very tedious task to keep a check on the quality of every sample produced in every nook and corner across the globe. In such a situation efficient result produced by time-saving and less labour-intensive ways like adapting machine learning techniques proves to be very helpful. Firstly, we applied a regression method to make metric evaluations and then secondly, we built an ANN model using the data set. We realized that the neural network model helped in attaining a better accuracy score of 0.

In the future, we can make use of other deep learning algorithms and techniques to attain better results. We can focus on other dimensional reduction techniques to yield better results. In addition to that, other advanced machine - learning techniques can be implemented for the classification of milk quality.

REFERENCES

[1] J. K. Swathi &N. Kauser (2015). A study on adulteration of milk and milk products from local vendors. International Journal of Biomedicaland Advance Research, 6(09), 678-681.

[2] S. Goyal, &G. K. Goyal, (2012). Artificial Neural Networks for Dairy Industry: A. *Journal of Advanced Computer Science and Technology*, *1*(3), 101-115.

[3] B. Vallejo- Cordoba, G. E. Arteaga, &S. Nakai, (1995). Predicting milk shelf- life based on artificial neural networks and headspace gas chromatographic data. *Journal of Food Science*, *60*(5), 885-888.

[4] O. Polat, S. G. Akçok, M. A. Akbay, D. Topaloğlu, S. Arslan, &C. B. Kalayci, (2021). Classification of raw cow milk using information fusion framework. *Journal of Food Measurement and Characterization*, *15*(6), 5113-5130.

[5] P. S. Vasafi, &B. Hitzmann, (2022). Comparison of various classification techniques for supervision of milk processing. *Engineering in Life Sciences*, *22*(3-4), 279-287.

[6] W. Habsari, F. Udin, & Y. Arkeman, (2022, July). An analysis and design of fresh milk smart grading system based on internet of things. In *IOP Conference Series: Earth and Environmental Science* (Vol. 1063, No. 1, p. 012059). IOP Publishing.

[7] S. Chaturvedi, A. K. Gupta, R. L. Yadav, & A. K. Sharma, (2013). Life time milk amount prediction in dairy cows using artificial neural networks. *International Journal of Recent Research and Review*, *5*, 1-6.

[8] P. Kominakis, Z. Abas, I. Maltaris, &E. Rogdakis, (2002). A preliminary study of the application of artificial neural networks to prediction of milk yield in dairy sheep. *Computers and electronics inagriculture*, *35*(1), 35-48.

[9] D. Banik, &R. R. Chowdhury, (2016, December). Offline signature authentication: A back propagation-neural network approach. In *2016 Sixth International Symposium on Embedded Computing and System Design (ISED)* (pp. 334-339). IEEE.

[10] B. Anuradha, & V. V. Reddy, (2008). ANN for classification of cardiac arrhythmias. *ARPN Journal of Engineering and Applied Sciences*, *3*(3), 1-6.

[11] D. Pujari, R. Yakkundimath, &A. S. Byadgi, (2016). SVM and ANN based classification of plant diseases using feature reduction technique. *IJIMAI*, *3*(7), 6-14.

[12] N. A. Mahmon & N. Ya'acob, (2014, August). A review on classification of satellite image using Artificial Neural Network (ANN). In *2014 IEEE 5th Control and system graduate research colloquium* (pp. 153-157). IEEE.

[13] G. Guo, H. Wang, D. Bell, Y. Bi, & K. Greer, (2003, November). KNN model-based approach in classification. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"* (pp. 986-996). Springer, Berlin, Heidelberg.

[14] S. Chandra, M. K. Gourisaria, G. M. Harshvardhan, S. S. Rautaray, M. Pandey, & S. N. Mohanty, (2021). Semantic Analysis of Sentiments through Web-Mined Twitter Corpus. In *ISIC* (pp. 122-135).

[15] A. Sahu, H, GM, M.K. Gourisaria, S. S. Rautaray, and M. Pandey, "Cardiovascular risk assessment using data mining inferencing and feature engineering techniques," *International Journal of Information Technology*, vol. 13, no. 5, pp. 2011–2023, Apr. 2021.

[16] M. K. Gourisaria, H, GM,, R. Agrawal, S. S. Patra, S. S. Rautaray, and M. Pandey, "Arrhythmia Detection Using Deep Belief Network Extracted Features From ECG Signals," *International Journal of E-Health and Medical Communications*, vol. 12, no. 6, pp. 1–24, Nov. 2021.

[17] S. Sarah, M. K. Gourisaria, S. Khare, &H. Das, (2022). Heart Disease Prediction Using Core Machine Learning Techniques—A Comparative Study. In *Advances in Data and Information Sciences* (pp. 247-260). Springer, Singapore.

[18] V. Singh, M. K. Gourisaria, & H. Das, (2021, September). Performance Analysis of Machine Learning Algorithms for Prediction of Liver Disease. In *2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON)* (pp. 1-7). IEEE.

[19] B. Vallejo- Cordoba, G. E. Arteaga, & S. Nakai, (1995). Predicting milk shelf- life based on artificial neural networks and headspace gas chromatographic data. *Journal of Food Science*, *60*(5), 885-888.