Research Article

Gaussian Naïve Bayes Algorithm: A Reliable Technique Involved in the Assortment of the Segregation in Cancer

M. Vijay Anand,[1] B. KiranBala,[2] S. R. Srividhya [ID],[3] Kavitha C. [ID],[3] Mohammed Younus,[4] and Md Habibur Rahman [ID][5]

[1]Department of Computer Science Engineering, Saveetha Engineering College, Chennai, India
[2]Department of Artificial Intelligence and Data Science, K.Ramakrishnan College of Engineering, Trichy, Tamil Nadu, India
[3]Department of Computer Science Engineering, Sathyabama Institute Science and Technology, Chennai, India
[4]College of Applied Computer Sciences, King Saud University, Al-Muzahmiyah Branch, Saudi Arabia
[5]Dept. of Computer Science and Engineering, Islamic University, Kushtia-7003, Bangladesh

Correspondence should be addressed to Md Habibur Rahman; habib@iu.ac.bd

Cancer is a disease caused by uncontrollable cell growth. The disease is a constant subject of concern due to unavailability of treatment at a severe level. Patients who have suffered from the disease have the chance of getting saved if this fatal illness is identified in the beginning stage. The survival chance will be very low if it is detected in the final stage of cancer. As the patients could not survive in their last stage, to cure their disease, an early diagnosis is a key issue and is vital. For the classification of cancer, Gaussian Naïve Bayes is implemented in this work. By exerting it on two datasets, the algorithm is tested, in which the Wisconsin Breast Cancer Dataset (WBCD) is considered as earliest one and the next one is the Lung Cancer Dataset. The assessment result of the suggested algorithm attained 90% accuracy in the prediction of lung cancer, and in predicting breast cancer, the accuracy is 98%.

1. Introduction

Classification is a task of searching or mining data that classifies items in an accumulation to the target classes [1]. In data, where the classes are prebuilt, the classification predicts each instance's target class and it is a supervised technique of learning [2]. The objective of classification is, that for every status in the data, the target class should be predicted accurately. Classification helps in the discovery of various models of concealed data that have been not transparent to the data annotator [3]. Medical diagnosis, detection of faults in industry, and recognition of patterns and images are some examples of applications of classification. It is the main side that helps us to select and make decisions [4] between two or more groups. The two parameters which are used to find the classification model's coherence are accuracy and interpretability. The disease of

cancer in many cases brings about the death of the majority of people who suffer from it [5]. Cancer is caused by the unusual extension of cells in any parts or tissues of the body. It may take place in any part of the body. We might be thinking that all types of tumors are not cancerous but the fact is, yes. Since it is basically unnecessary, some tumor types do not unfold to all parts of the body. The different indicators of cancer include tumor, an abnormal type of bleeding, cough for a considerably long time, and loss of weight. Earlier recognition [6] is the most productive way to abate the deaths of cancer. Without going for surgical operations, early detection needs a well precise and dependable procedure of detection that permits doctors to separate benignant lumps from cancerous ones [7]. Altogether, there are more or less than 100 kinds of cancers that affect the body of human beings. Some kinds of cancers are as follows:

(a) Of the two datasets, lung cancer is the anomalous cell's unusual growth that undertakes in a single lung or both of the lungs usually in the form of cells lining the passages of air [8]. The healthy tissues can be destructed by the growth of abnormal cells; they divide quickly and the formation of tumor occurs. Basically, cancer in the lungs is divided into two. They are nonsmall cell lung carcinoma (NSCLC) and small cell lung carcinoma (SCLC). Furthermore, NSCLC is broadly divided into adenocarcinoma, large cell carcinoma, and squamous cell carcinoma. SCLC is the second type. There are also various histologic groupings in SCLC which include large cell carcinoma, combined small cells, pure small cells, and also mixed small cells. The more attacker than NSCLC is deadlier than SCLC [9]

(b) The other dataset describes that cancer which affects females worldwide is breast cancer [10]. It remains to be one of the most popular cancers in the world. In 2017, from all of the cancer status, the estimation seems to be 25% that detected amidst 1.67 million statuses of cancer were breast cancer. Specifically, compared to the developed nations, the women from less developed nations have a diminutive number of cancer statuses, like 794000 in advanced nations and 883000 cancer statuses in less advanced nations [11]. In comparison, the breast cancer rate in the United Kingdom (UK) is higher than the case of India where the rate is 28.8 per 100000, while in the case of UK, the rate is 95/100000, but the rate of death equals (12.7 vs. 17.1/100000) that of the UK.

There are several techniques of classification in literature, which are familiar with the detection of cancer like the genetic algorithm, C4.5, and using support vector machines (SVMs) [12]. Gaussian Naïve Bayes is implemented in this study to classify cancer that aids doctors to diagnose this deadly infection in the early stage and to keep the patient's life [13].

2. Related Works

Data mining techniques are used by several researchers for the detection of various diseases. Jyothy Sony deployed classifiers such as K-nearest neighbors [14] and Naïve Bayes. Classifiers have been used to datasets of Wisconsin Prognostic Breast Cancer (WPBC) [15] in Naïve Bayes of Diana Dumitru's research, concerning a binary class: instance's nonrecurrent events and 47 instances' recurrent events and also several 198 patients. Mohd Fauzi et al. presented different classification techniques' comparisons such as radial basis, Bayes network, and pruned tree. On a large dataset, for knowledge analysis, the K-nearest neighbors algorithm uses Waikato to Environment. Breast cancer is the data used in their investigation. Alazab et al. [16] analyzed the supervised learning algorithm's performance such as SVM Gaussian RBF Kernel, RBF neural networks, simple CART, decision tree, and Naïve Bayes. These algorithms are accustomed to classifying the datasets of breast cancer [17]. Shandilya and

Chandankhede designed the system of an intelligent and effective prediction of heart disease using the associative classifier's weighted concept [18]. As compared to the other already existing associative classifiers, it gives better accuracy. The accuracy was found to be 81.51% for weighted associative classifiers (WAC). According to their capability in predicting, WAC assigns different weights to different attributes. With regards to accuracy [19], experiments are being conducted on a dataset of benchmark to assess the performance [20].

3. Naïve Bayes

It is an algorithm of effortless learning that utilizes the rule of Bayes in along with a high premise or assumption whose characteristics are depending on the independence provided by the class. Though Naïve Bayes brings the accuracy of competitive classification again and again, while assuming this independently often invaded in practice [21]. Practically, the combination of computational efficiency and many other advantageous aspects may lead to naïve Bayes being largely applied. Evaluate the posterior probability value of $P(y|x)$, each of the class of y, with an object x. Using the details inside the sample data, Naïve Bayes contributes an apparatus. If such evaluations are done once, they can be used for the classifications or any other application of decision support [22].

The following are the features of Naïve Bayes:

(i) Low variance

As the search is not utilized by Gaussian Naïve Bayes, it contains variance at a low value, despite the cost of bias being high

(ii) Incremental learning

In general, the Gaussian Naïve Bayes functions from probabilities of the lower-order estimates obtained from that of the training data. As new training data are obtained, these can be updated readily.

(iii) Computational efficiency

The training time equals the attributes and training examples, while the value of classification time equals the attributes and it is unaffected by the attributes number.

(iv) Robustness in the face of absent value

If one of the attribute values is missing and information from another attribute is still used, then it results in the sophisticated deterioration in performance, as for all predictions, all attributes are always used by the naïve Bayes. Due to its probabilistic framework, it is extremely thoughtless of the value of missing attributes in the data deployed in the training process.

(v) Robustness in facing noise

For all predictions, Gaussian Naïve Bayes uses the whole attributes. Hence, it is verbally thoughtless to noise in the examples which is needed to be assorted and also for noising in the training data as it uses probabilities.

TABLE 1: Naïve Bayes classifier's accuracy of classification (by using a dataset with preprocessing and without preprocessing).

Classified instances	Correctly	Incorrect	Precision	Recall	F-measure
Without preprocessing and feature selection	2713 (96.9%)	87 (3.1%)	0.969	0.969	0.969
With preprocessing and feature selection	2675 (95.5%)	125 (4.5%)	0.956	0.955	0.955

4. Naive Bayes Classifier

Learning can be greatly simplified by the Naïve Bayes classifier by supposing that features are independent given class [23]. Although, the assumptions of independence are poor in general. Practically, with a more sophisticated classifier, Naive Bayes often competes effectively. The feature vector described that to a given example, Bayesian classifiers allocate the most likely class, i.e.,

$$P(C) = \prod_{i}^{n} = 1 p(X_i | C), \tag{1}$$

where C is the classifier, and $X = (X_1, \ldots \ldots \ldots, X_n)$ is a feature vector [24].

In practice, the resulting classifier known as Naïve Bayes is very successful despite this unrealistic assumption, by competing again and again with much more sophisticated techniques. The model of Naïve Bayes is a simplified version of Bayesian probability. The operations of the Naïve Bayes classifier are done on a sturdy independence assumption. This means that one attribute of the probability of one does not have any effect on the probability of the rest of the attributes [25]

For the feature selection, rank search and Cfs subset evaluator are used. Therefore, the adoption of another technique of rank search has been tried. 89 attributes are selected, and chi-square feature selection has been adopted this time. 89 attributes have been inputted this time in preference to the previously selected 75 attributes. After using feature selection of chi-square, the result has been improved, 0.1% of improved accuracy. Even though there is an insignificant improvement, it is certified that to achieve better classification results, feature selection and preprocessing are really useful, as given in Table 1. It is believed that under different situations, to fulfill different classification results, different searching techniques can help. To generalize the best solution, there are needs to take many trials and time. However, due to the constraints of time, this study concludes only by using feature selection and preprocessing which can attain better classification outcomes [26, 27].

5. Cancer Classification

Clinical prediction of results in cancer is generally accomplished by the evaluation of histopathology of samples of the tissue procured during primary tumor surgical removal. The characteristics of histopathology include tissue integrity, the size of the tumor, histological grade, a typical cell morphology, irregular expression of genetic markers and protein, malignant transformation's evidence, proliferation and senescence, invasion depth, invasive margin's characteristics (IM), and the extent of vascularization [28]. By the system of classification, the progression of cancer's evaluation is carried out longitudinally and then put in to estimate patient prognosis. It is important to begin the assimilation of the immunocore as a cancer classification component as it is considered the imperative role of the host immune signature to control the progression of tumor and a prognostic tool. The Gaussian Naive Bayes algorithm is shown in Algorithm 1. There are two advantages to this strategy.

(1) It seems to be the strapping symptom in the case of overall survival [29, 30] (OS) as well as disease-free survival (DFS), especially in the first stage of cancer

(2) For novel therapeutic approaches, it could suggest potential goals including immunotherapy; concerned with procedure of prognostic and diagnostic analysis of tumors, the technologies of ongoing immunohistochemical allow the application of such assessments by empirical labs.

6. Proposed Method

We can see the scheduled work's main phases in this section (classification of cancer by using Naïve Bayes) as shown in Figure 1.

6.1. Data Acquisition Stage. From cancer, the classification proposed was executed in two sets.

For classification, from the machine learning bin of UCI, the Wisconsin Breast Cancer is quoted as a standard dataset where the database contains 699 instances and 11 attributes are the first type. The dataset of Lung Cancer which is procured from the original website where the database contains 23 attributes and 1000 instances of data (expecting ID number and class label) is the second type. It describes gender, age, pollution of air, allergy to dust, use of alcohol, genetic risk, occupational hazards, obesity, balanced diet, chronic lung disease, passive smoker, smoking, coughing of blood, chest pain, fatigue, shortness of breath, weight loss, snoring, frequent cold, swallowing difficulty, dry cough, wheezing, and clubbing of fingernails. The attributes of the Breast Cancer dataset are given in Table 2.

6.2. Data Preprocessing. In data mining, data preprocessing acts as the initial step. Before exerting data preprocessing for any type of classification, the database that is made useful in research must be preprocessed. This dataset may hold on data and leads up to deceiving results. Therefore, the data's quality should be improved before working on an evaluation. Data normalization is the data preprocessing done for this work. The technique which was used to sniff out the attribute values that are worthy of being zero and are not

Input: dataset
      S//dataset attributes involved
Output: class name
      Start
(1) To train the date splitting the dataset into 70%, the remaining 30% is used for testing.
(2) Training phase
(2.1)   Total = all instances in the training dataset
             C_j is class in the dataset of training.
(2.2)   Probability of every single class is calculating
             $P(C_j)$ = frequency (C_j)/total
(2.3)   Calculate the mean (μ) as well as the standard deviation (σ) values of each of the training dataset class attributes. Note down the result.
(3) Testing phase
(3.1)   In testing DSX is an instance.
(3.2) By applying equation (1), the probability density function (PDF) value of X is calculated at C_p for values of X attributes remain in S, $p(X_i|C_j)$.
(3.3) By using the equation, $P(X|C_j) = \prod_{j=1}^{n} P(f_j|C_i)$, for the values, resulting from step of 3.2, the conditional probability value of X is calculated at C_j,
(3.4) By using an equation, $P(X) = P(c_j).p(C_j)$, here, $p(C_j|X)$ represents the probability value of instance at C_j, and then, the posterior probability of X can be calculated.
(3.5) By selecting maximization $P(C_j|X)$, assign a X class label.
(4) Return the class name
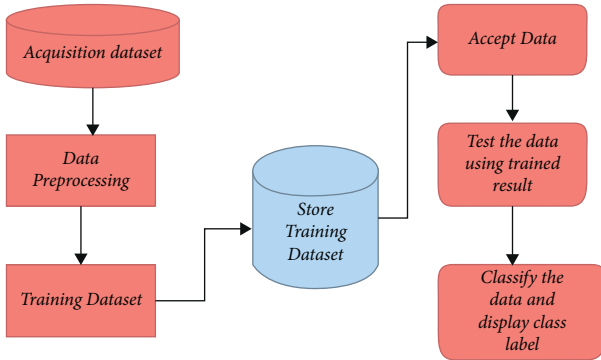   End of the process

ALGORITHM 1: Gaussian Naïve Bayes.

FIGURE 1: Proposed system' block diagram.

effectual in this procedure of classification is the z-score normalization technique.

7. Results and Discussion

7.1. Accuracy. Accuracy discloses, in general, how the model works and whether a model is being trained properly and correctly. However, it does not give specific details on how it tried to the problem. As your primary success, the demerit in using machine learning (ML) accuracy is that, when there is a significant class divide, it fails inadequately. Accuracy may be false for the unbalanced datasets. For evaluating classification models, accuracy is one metric. It has the following definitions formally:

$$\text{Accuracy} = \frac{\text{No. of correct predictions}}{\text{Total no. of predictions}}. \tag{2}$$

Accuracy is calculated with the positives and negatives for binary classification as follows:

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FN + FP}. \tag{3}$$

where TP is the value of true +ive; FP is the value of False +ive; TN is the value of true −ive; and FN is the value of false −ive.

7.2. Precision. When cost of false positives (FP) is high, then the precision helps. So, let us assume that the issue involves skin cancer detection. Many of the patients will tell that they have melanoma if we have a very low precision model and that may include some false detection. After being assaulted with false alarms, those who detect the results will ignore them when the false positives are too high.

$$\text{Precision} = \frac{TP}{TP + FP}. \tag{4}$$

7.3. TPR. TPR (true positive rate), which is also called recall, is metric related to the classification model. It is calculated as

$$TPR = \frac{TP}{TP + FN}. \tag{5}$$

The probability that an actual positive will test positive is true positive rate.

Table 2: Breast Cancer dataset's attributes.

Attribute	Domain
ID (code number 1–10)	
Uniformity of cell shape	
Single epithelial cell size	
Mitosis	
Marginal adhesion	
Uniformity of cell size	
Normal nucleoli	
Clump thickness	
Bland chromatin	
Bare nuclei	
Class	Benign, 2
	Malignant, 4

7.4. DSC.

DSC (dice similarity coefficient) is also known as Sorensen Dice Index. To measure the similarity of two samples, DSC is used and it is a statistic matrix.

By using the definition of the value of true +ive (TP), false −ive (FN), and false +ive (FP), when applied to Boolean data, it can be written in the form of

$$\frac{2TP}{2TP + FP + FN}. \tag{6}$$

For the implementation of algorithms, the Visual Basic Net.2013 is viable [29]. We have executed assessments on datasets of breast cancer which consists of 699 instances, of which 241 instances are huge and 458 instances are benignant and also applied on datasets of lung cancer containing thousand instances, of which more than 300 records of the file are of lower level, 332 records of the file are of the medium level, and 365 high-level records for the effectiveness of evaluation of the proposed algorithm. For learning and training, each dataset is split into 70% and for testing, 30%. For the first dataset, it is benign or malignant to segregate the data. For the second dataset, categorize the data irrespective of its level such as low, medium, or high level. Three metrics have been calculated, for assessing the execution system proposed. These metrics and their formulas are

$$\text{Accuracy} = \left(\frac{TN + TP}{\text{All}}\right) * 100\%,$$

$$\text{Specificity} = \left(\frac{TN}{FP + TN}\right) * 100\%, \tag{7}$$

$$\text{Sensitivity} = \left(\frac{TP}{FN + TP}\right) * 100\%.$$

where TP indicates true +ive and TN indicates true −ive and FN and FP indicated false +ive and false −ive, respectively.

A confusion matrix is a basically a table of data that holds the details about predicted and actual classification systems attained by a system of classification. Data used for the evaluating the execution of such a system are given in Table 3.

Notwithstanding, to compute specificity, accuracy, and sensitivity for both datasets, these values of metrics were

Table 3: Two classes confusion matrix.

		Predicted	
		−ive	+ive
Actual	−ive	TN	TP
	+ive	FN	FP

Table 4: Breast cancer's confusion matrix values.

		Predicted class	
		Negative	Positive
Actual	Negative	74	1
	Positive	3	137

Table 5: Lung cancer's confusion matrix values.

		Predicted class level		
		High	Medium	Low
Actual class level	High	83	3	7
	Medium	0	107	3
	Low	0	17	85

Table 6: Naïve Bayes classifier's performance evaluation of Gaussian Naïve Bayes.

Type of cancer dataset	Sensitivity (%)	Accuracy (%)	Specificity (%)
Breast	97.5	97.5	98.1
Lung	92	91	96.5

Table 7: Proposed algorithms and another previously existing algorithm comparison.

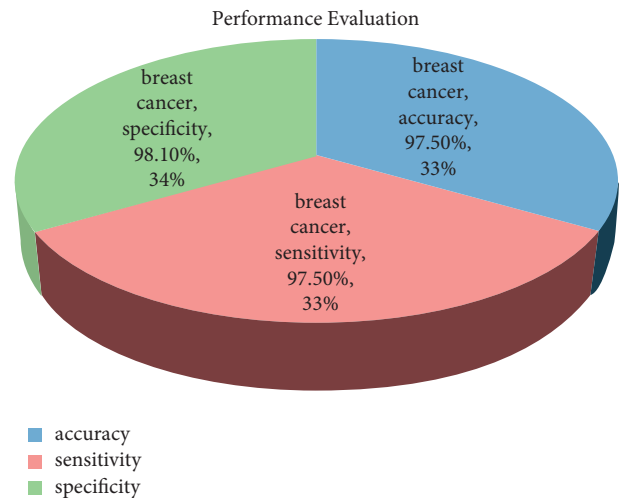| Algorithm	Proposed algorithm (%)	RBF network (%)	J48 (%)
Accuracy	98.5	96.7	94

Figure 2: The output of the performance evaluation of cancer (breast).

used, as given in Tables 4 and 5, which indicate the matrix of confusion for the preprocessed system. Table 6 provides the result of the Naïve Bayes classifier's performance evaluation of Gaussian Naïve Bayes.

In cancer classification, empirical effects display that the algorithm was the best as shown in the other previously existing algorithm comparison in Table 7 and Figure 2.

8. Conclusion

To identify the disorganized value in the classification for normalization, the proposed work presents an effective approach, which deserves to be zero. For the prediction of cancer based upon Gaussian distribution, the algorithm of Naïve Bayes is used. Among the two types of cancer named lung cancer as well as breast cancer, the model was implemented, and to list the patient's details, an interface has been intended, which predicts the class of cancer. Therefore, high classification performance has been achieved by the proposed work. In this study, the author's is early detection of cancer using the Naïve Bayes algorithm. The survey presented in this study shows different cancer classifications and found that Naïve Bayes is an algorithm of simple learning and learned about Naïve Bayes classifiers. This research helps in the success of the project.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] M. Amrane, S. Oukid, I. Gagaoua, and T. Ensari, "Breast cancer classification using machine learning," in *Proceedings of the 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, pp. 1–4, Istanbul, Turkey, April 2018.

[2] K. Arutchelvan and R. Periasamy, "Analysis of cancer detection system using data mining approach," *Int J Innov Res Adv Eng*, vol. 11, no. 2, pp. 57–60, 2015.

[3] H. Asri, H. Mousannif, H. A. Moatassime, and T. Noel, "Using machine learning algorithms for breast cancer risk prediction and diagnosis," *Procedia Computer Science*, vol. 83, pp. 1064–1069, 2016.

[4] D. Berrar, "Bayes' theorem and naive Bayes classifier," *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, pp. 403–412, Elsevier Science Publisher, Amsterdam, The Netherlands, 2018.

[5] P. Bethapudi, E. SreenivasaReddy, and T. Sitamahalakshmi, "A computational approach for better classification of breast cancer using GeneticAlgorithm," *International Journal of Engineering and Computer Science*, vol. 4, 2015.

[6] A. Bharat, N. Pooja, and R. A. Reddy, "Using machine learning algorithms for breast cancer risk prediction and diagnosis," in *Proceedings of the 2018 3rd International Conference on Circuits, Control, Communication and Computing*, no. I4C, pp. 1–4, Bangalore, India, October 2018.

[7] A. A. Farid, G. Selim, and H. Khater, "Applying artificial intelligence techniques for prediction of neurodegenerative disorders," *A comparative case-study on clinical tests and neuroimaging tests with Alzheimer's Disease*, 2020.

[8] J. Galon, F. Pagès, F. M. Marincola et al., "Cancer classification using the Immunoscore: a worldwide task force," *Journal of Translational Medicine*, vol. 10, no. 1, p. 205, 2012.

[9] G. Kesavaraj and S. Sukumaran, "A Study on Classification Techniques in Data Mining. 2013 Fourth International Conference on Computing," in *Proceedings of the Communications and Networking Technologies (ICCCNT)*, pp. 1–7, Tiruchengode, India, July 2013.

[10] S. Kharya, S. Agrawal, and S. Soni, "Naive Bayes classifiers: a probabilistic detection model for breast cancer," *International Journal of Computer Application*, vol. 92, no. 10, pp. 26–31, 2014.

[11] G. T. Reddy, M. P. K. Reddy, K. Lakshmanna, D. S. Rajput, R. Kaluri, and G. Srivastava, "Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis," *Evol. Intel.*vol. 13, no. 2, pp. 185–196, 2020.

[12] S. Kharya and S. Soni, "Weighted naive bayes classifier: a predictive model for breast cancer detection," *International Journal of Computer Application*, vol. 133, no. 9, pp. 32–37, 2016.

[13] F. malar and D. mbaiya, "Application of data mining techniques in early detection of breast cancer," *International Journal of Engineering Trends and Technology*, vol. 56, no. 1, pp. 43–45, 2018.

[14] D. D. Lewis, "Naive (Bayes) at forty: the independence assumption in information retrieval," *European Conference on Machine Learning*, vol. 1398, pp. 4–15, 1998.

[15] S. Mukherjee and N. Sharma, "Intrusion detection using naive Bayes classifier with feature reduction," *Procedia Technology*, vol. 4, pp. 119–128, 2012.

[16] M. Alazab, K. Lakshmanna, T. R. G, Q. V. Pham, and P. K. Reddy Maddikunta, "Multi-objective cluster head selection using fitness averaged rider optimization algorithm for IoT networks in smart cities," *Sustainable Energy Technologies and Assessments*, vol. 43, 2021.

[17] N. M. Mule, D. D. Patil, and M. Kaur, "A comprehensive survey on investigation techniques of exhaled breath (EB) for diagnosis of diseases in human body," *Informatics in Medicine Unlocked*, vol. 26, p. 100715, Article ID 100715, 2021.

[18] S. Shandilya and C. Chandankhede, "Survey on Recent Cancer Classification Systems for Cancer Diagnosis," in *Proceedings of the 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pp. 2590–2594, Chennai, India, March 2017.

[19] C. Iwendi, P. K. R. Maddikunta, T. R. Gadekallu, K. Lakshmanna, A. K. Bashir, and M. J. Piran, "A meta-heuristic optimization approach for energy efficiency in the IoT networks," *Software: Practice and Experience*, vol. 51, no. 12, pp. 2558–2571, 2021.

[20] S. Ting, W. Ip, and A. H. Tsang, "Is Naive Bayes a good classifier for document classification," *International Journal of Software Engineering and Its Applications*, vol. 5, no. 3, pp. 37–46, 2011.

[21] T. R. Gadekallu, N. Khare, S. Bhattacharya et al., "Early detection of diabetic retinopathy using PCA-firefly based deep learning model," *Electronics*, vol. 9, no. 2, p. 274, 2020.

[22] D. Umesh and C. Thilak, "Predicting breast cancer survivability using Naïve Baysien and C5. 0 algorithm," *International Journal of Computer Science and Information Technology Research*, vol. 3, no. 2, pp. 802–807, 2015.

[23] T. R. Gadekallu, M. Alazab, R. Kaluri, P. K. R. Maddikunta, S. Bhattacharya, and K. Lakshmanna, "Hand gesture classification using a novel CNN-crow search algorithm," *Complex & Intelligent Systems*, vol. 7, no. 4, pp. 1855–1868, 2021.

[24] G. I. Webb, E. Keogh, and R. Miikkulainen, "Naïve bayes," *Encyclopedia of Machine Learning*, vol. 15, pp. 713-714, 2010.

[25] F.-J. Yang, "An implementation of naive bayes classifier," in *Proceedings of the 2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, pp. 301–306, Las Vegas, NV, USA, December 2018.

[26] S. Pandya, T. R. Gadekallu, P. K. Reddy, W. Wang, and M. Alazab, "InfusedHeart: a novel knowledge-infused learning framework for diagnosis of cardiovascular events," *IEEE Transactions on Computational Social Systems*, pp. 1–10, 2022.

[27] E. A. Mantey, C. Zhou, S. R. Srividhya, S. K. Jain, and B. Sundaravadivazhagan, "Integrated blockchain-deep learning approach for analyzing the electronic health records recommender system," *Frontiers in Public Health*, vol. 10, Article ID 905265, 2022.

[28] T. R. Gadekallu, N. Khare, S. Bhattacharya, S. Singh, P. K. R. Maddikunta, and G. Srivastava, "Deep neural networks to predict diabetic retinopathy," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–14, 2020.

[29] G. T. Reddy, M. P. K. Reddy, K. Lakshmanna et al., "Analysis of dimensionality reduction techniques on big data," *in IEEE Access*, vol. 8, pp. 54776–54788, 2020.

[30] C. Kavitha, V. Mani, S. R. Srividhya, O. I. Khalaf, and C. A. Tavera Romero, "Early-stage alzheimer's disease prediction using machine learning models," *Frontiers in Public Health*, vol. 10, Article ID 853294, 2022.