

An Ontological Model based on Machine Learning for Predicting Breast Cancer

Hakim El Massari^{1*}, Noredine Gherabi², Sajida Mhammedi³, Hamza Ghandi⁴, Fatima Qanouni⁵, Mohamed Bahaj⁶
National School of Applied Sciences, Sultan Moulay Slimane University, Lasti Laboratory, Khouribga, Morocco^{1, 2, 3, 4, 5}
Faculty of Sciences and Technologies, Hassan First University, Settat, Morocco⁶

Abstract—Breast cancer is mostly a female disease, but it may affect men as well even at a considerably lower percentage. An automated diagnosis system should be built for early detection because manual breast cancer diagnosis takes a long time. Doctors have lately achieved significant advances in the early identification and treatment of breast cancer in order to decrease the rate of mortality caused by the latter. Researchers, on the other hand, are analysing large amounts of complicated medical data by employing a combination of statistical and machine learning methodologies to assist clinicians in predicting breast cancer. Various machine learning approaches, including ontology-based Machine Learning methods, have lately played an essential role in medical science by building an automated system that can identify breast cancer. This study examines and evaluates the most popular machine learning algorithms, besides the ontological model based on Machine Learning. Among the classification methods investigated were Naive Bayes, Decision Tree, Logistic Regression, Support Vector Machine, Artificial Neural Network, Random Forest, and k-Nearest Neighbours. The dataset utilized has 683 instances and is available for download from the Kaggle website. The findings are assessed using performance measures generated from the confusion matrix, such as F-Measure, Accuracy, Precision, and Recall. The ontology model surpassed all machine learning techniques, according to the results.

Keywords—Machine learning; prediction; ontology; semantic web rule language; decision tree; breast cancer

I. INTRODUCTION

In 2020, 2.3M women were identified with breast cancer, with 685 thousand fatalities worldwide. By the end of 2020, there will have been 7.8M women diagnosed with breast cancer in the previous five years, making it the most frequent kind of cancer in the world. Breast cancer claims more DALYs from women than any other kind of cancer worldwide. Breast cancer affects women of all ages after puberty in every country throughout the world, but at a rising rate in the latter stages of life. From the 1930s through the 1970s, there was minimal change in the breast cancer death rate. In nations where early diagnosis systems are available in conjunction with various treatment options to remove intrusive illnesses, life expectancies started to get better as in 1980s.

Machine learning (ML) is one of the most constantly evolving areas of computer science, with a wide range of applications [1], [2]. It is the process of obtaining usable information from a big quantity of data [3]. Marketing, Industry, Medical diagnosis, and other scientific domains all make use of ML approaches. ML algorithms are well-suited for

medical data analysis since they have been frequently employed in medical datasets. ML comes in several forms, including classification, regression, and clustering. Each form has a particular consequence and influence depending on the problem that we are attempting to address. We focus on classification algorithms in our work because of their high accuracy and performance in classifying a given dataset into predetermined categories and predicting future events or information from that data. In the medical field, classification algorithms are often utilized, particularly in the diagnosis of illnesses such as breast cancer. As a result, regularly used machine learning classification methods such as Support Vector Machine (SVM), k-Nearest Neighbours (KNN), Artificial Neural Network (ANN), Naive Bayes (NB), Logistic regression (LR), and Decision Tree (DT) are utilized to detect patients with breast cancer at an early stage.

Several researches have been conducted, and various machine learning models have been implemented, to identify and predict breast cancer diagnoses [4]. For example, this study [5] sought to identify the most accurate machine learning approaches for predicting patients with breast cancer. Various supervised machine-learning algorithms, including RF, KNN, DT, AdaboostM1, LR, and ANN [6], were used and their performance was compared. The authors of this study [7], examined linear discriminant analysis with support vector machines in terms of specificity, F-Measure, accuracy, and sensitivity to see which method is better for classifying breast cancer datasets. The results reveal that the support vector machine outperforms the linear discriminant analysis, with a 98.77 percent accuracy rate. [8] This study covers the whole Bayesian technique to assess the predicted distribution of all classes using three datasets and three classifiers: naive bayes (NB), bayesian networks (BN), and tree augmented naive bayes (TAN). The outcomes showed that the BN method performed the best, with an accuracy rate of 97 percent.

Breast cancer diagnosis and prediction has garnered a lot of attention in recent years, and numerous ways have been taken to address this issue [9]–[12]. The present focus is on machine learning and the semantic web. The Wisconsin Breast Cancer Dataset was used in this study [13]. The authors' purpose is to study the dataset and assess the efficacy of several ML algorithms for predicting breast cancer. Several machine learning methods have been developed to differentiate benign and malignant tumors. To predict whether breast cancer is malignant or benign, [14] the authors used machine learning and computer vision to extract features and construct an optimized model by varying hyper-parameter values. The

*Corresponding Author.

analysis is carried out using a support vector machine, and several quality indicators are produced, with the observed findings being notable. [15] The goal of this study is to evaluate the classification algorithms' prediction accuracy in terms of efficiency and effectiveness. The authors conduct a rigorous comparison of classification algorithms such as SVM, DT, NB, and RF in terms of prediction accuracy utilizing WEKA and a 10-fold cross validation approach on the Wisconsin Diagnostic Breast Cancer dataset. In this study [16], two machine learning algorithms, Decision Tree Classifier, and Logistic Regression, were implemented for breast cancer prediction on the "Breast Cancer Wisconsin (Diagnostic) Data Set," and their accuracies were compared and the Decision Tree Classifier is the most suited-algorithm for prediction since it has a precise prediction accuracy.

Recently, researchers published a significant quantity of research utilizing machine-learning algorithms to diagnose breast cancer [17]–[20]. In this comparison study [21], 4 machine learning (ML) algorithms were used: DT classifiers, SVM, KNN, and RF, and the results show that Support Vector Machine has the highest accuracy of 97% among them for the classification of breast tumors in women. On the Wisconsin Breast Cancer dataset, [22] the authors examined five supervised machine learning algorithms: LR, RF, KNN, ANN, and SVM. The metrics of the confusion matrix are used to assess the study's performance. According to the results, the ANNs had the greatest accuracy score of 98.57 percent.

Furthermore, ontology has been one of the most widely used techniques to managing, organizing, and extracting data

throughout the last few decades. It is a way of data representation that has been effectively utilized in a number of domains, particularly the medical domain. It is significant in computer science because of its ability to express many concepts and their relationships across fields. In reality, no single ontology is sufficient to meet today's expanding healthcare demands, and ontologies must be combined with machine learning algorithms to facilitate data integration and analysis. The authors in [23] created and explored an ontology-based decision tree model able to predict diabetes, [24] then compared the findings to numerous ML techniques, and discovered that the ontology model outperforms all other classifiers.

In this research, we intend to compare seven prominent classification approaches with the Ontological Model using carefully chosen criteria obtained from the confusion matrix, such as F-Measure, Accuracy, Precision, and Recall. The rest of this paper is organized as follows: Section II describes the methodologies utilized in this comparison analysis. Section III summarizes the findings and discussion. Section IV concludes and discusses future work.

II. METHODS AND EVALUATION

The approaches and materials employed, as well as the experimental methodology, dataset description, machine learning algorithms, ontology model, and evaluation metrics, are all included in this section. Fig. 1 depicts the process flowchart for this comparative study.

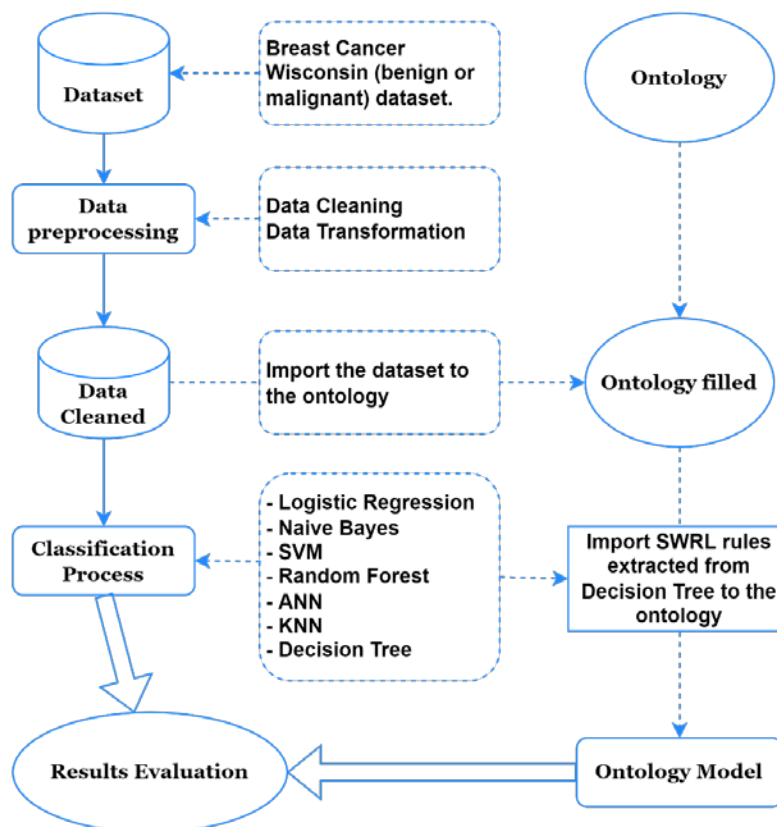


Fig. 1. Experimental Workflow.

A. Data Preprocessing

The dataset used is Breast Cancer Wisconsin - benign or malignant from Kaggle website, it consists of 683 instances and 10 features (9 attributes and the last one is a target). A full description of all dataset attributes is provided in Table I.

To build an effective machine learning classifier, we should always start with data cleaning, normalization of features, transformation of features, and even creation of new features from the dataset. The dataset contains 234 similar instances, after removing duplicated instances the remaining is 449 instances, where 213 represent benign cancer cells and 236 represent malignant cancer cells. We would like to inform you that in order to provide a fair comparison of the classification results obtained, we did not use any feature selection or performance-boosting methods.

B. Machine Learning Algorithms

We have used Weka software for all machine learning algorithms to predict whether the cancer cells are benign or malignant. Weka comprises tools for data classification, clustering, visualization, preparation, association rules mining, and regression [25].

We used the seven most classifiers used to classify datasets (Decision Tree, Random Forest, Logistic Regression, Artificial Neural Network, Naïve Bayes, Support Vector Machine, k-Nearest Neighbours). In addition, we employed two modes of test options: 10-fold cross validation and percentage split (split 50% train, remainder test) for the reason of enriching the study.

C. Ontological Model

This section presents the technologies used to create the ontology, besides the approach used to build the ontology model with the help of rules extracted from DT. This methodology has been referred to in this research for more details [26], which we recommend reading for more information. We'll go through some specifics shortly here.

1) *Ontology construction*: The ontology was built using the Protégé software, which is an open-source platform that provides a set of tools to a growing user community for constructing domain models and knowledge-based applications with ontologies. The ontology was created manually; the main classes are Diagnostic and Patient. The graphical representation of the ontology is shown in Fig. 2.

TABLE I. DATASET FEATURE'S INFORMATION

Attribute	Description
1- clump	Clump Thickness: Benign cells often form monolayers, whereas malignant cells frequently form multilayers.
2- uc2	Uniformity of Cell Size: Cancer cells differ in size.
3- uc3	Uniformity of Cell Shape: Cancer cells differ in form.
4- adhesion	Marginal Adhesion: Adhesion loss is an indication of cancer.
5- epithelial	Single Epithelial Cell Size: Is connected to the previously mentioned uniformity. Significantly expanded epithelial cells may be cancerous cells
6- bare_nuclei	Bare Nuclei: These are common in benign tumors.
7- bland_chromatin	Bland Chromatin: In benign cells, the nucleus has a homogenous texture.
8- normal_nucleoli	Normal Nucleoli: In normal cells, the nucleolus is generally quite tiny, if at all detectable. The nucleoli grow more visible in cancer cells.
9- mitoses	-
10- Class	Predicted class (2 for benign, 4 for malignant).

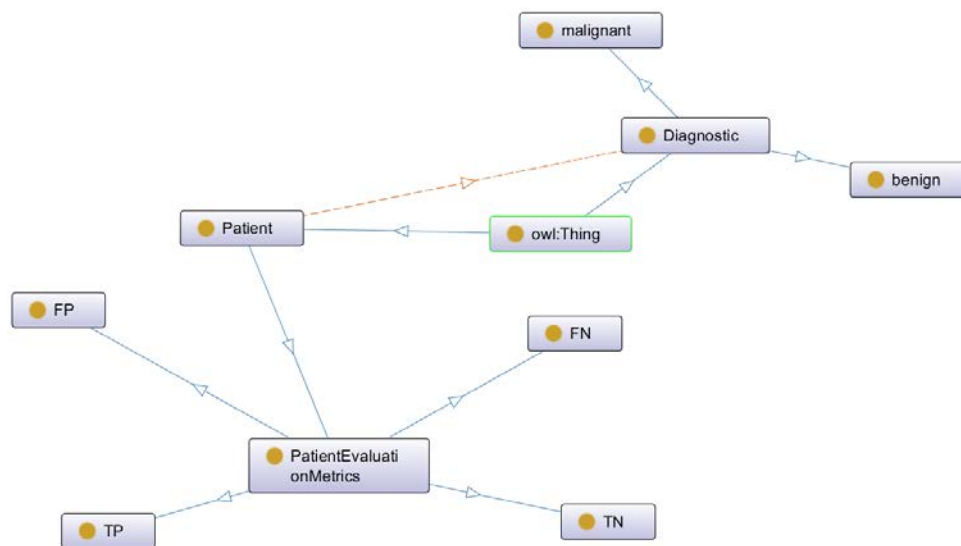


Fig. 2. The Ontology Graph.

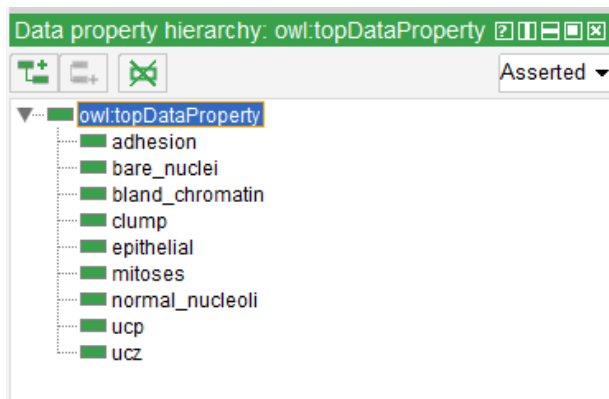


Fig. 3. Data Properties.

2) *Data properties and instances*: The data properties used in the ontology are the same attributes presented in Table I which are used to build models of machine learning algorithms. Fig. 3 illustrates the data properties. A plugin among the Protégé software plugins called Cellfie is used to import the same dataset used in Weka.

3) *Semantic web language rules and pellet reasoned*: Following the creation of classes, data properties, and instances in the ontology. We need to establish the SWRL reasoning rules. To achieve this, we used the SWRLTab plugin, we retrieved created rules from the DT algorithm, and imported them into Protégé. The collected rules from the DT algorithm are converted using the Java programming language, with each leaf of the tree extracted as a single SWRL rule. For instance/

A leaf from the DT algorithm

If ucp > 2 && ucZ ≤ 4 && bare_nuclei ≤ 2 && adhesion ≤ 3 THEN put the patient in benign

SWRL resulted

Patient(?P) ^ ucp(?P, ?UCP) ^ swrlb:greaterThan(?UCP, '2'^xsd:decimal) ^ ucZ(?P, ?UCZ) ^

swrlb:lessThanOrEqual(?UCZ, '4'^xsd:decimal) ^ bare_nuclei(?P, ?BN) ^ swrlb:lessThanOrEqual(?BN, '2'^xsd:decimal) ^ adhesion(?P, ?A) ^ swrlb:lessThanOrEqual(?A, '3'^xsd:decimal) → benign

To execute SWRL rules and infer new ontology axioms we utilized another plugin from Protégé software named Pellet [27], which includes capabilities for checking ontology coherence, deals with SWRL rules, computing the classification hierarchy, deals with OWL, explaining inferences, and answering SPARQL queries. It uses the Ontology and SWRL rules to initiate the inference and then determines if the cancer cells are benign or malignant. The ontology classifier's results are reported in the next section.

D. Evaluation

ROC Area, F-Measure, Root mean squared error, Recall, Accuracy, Root relative squared error, Precision, Kappa statistic, and other performance measures are employed to assess ML algorithms. We employed two test modes (split-test and K-fold cross-validation) using several metrics including Recall, F-Measure, Accuracy, and Precision to analyze our experimental results, which are presented below and in Fig. 4. Furthermore, the same criteria are utilized to assess the validity of this comparison research including ML classifiers and the ontological model.

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$PREC = \frac{TP}{TP+FP} \tag{2}$$

$$REC = \frac{TP}{TP+FN} \tag{3}$$

$$F\text{-Measure} = 2 * \frac{PREC*REC}{PREC+REC} \tag{4}$$

Other metrics, such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), are available but are most commonly employed in regression issues. As a result, owing to classification issues imposed by the dataset and techniques employed.

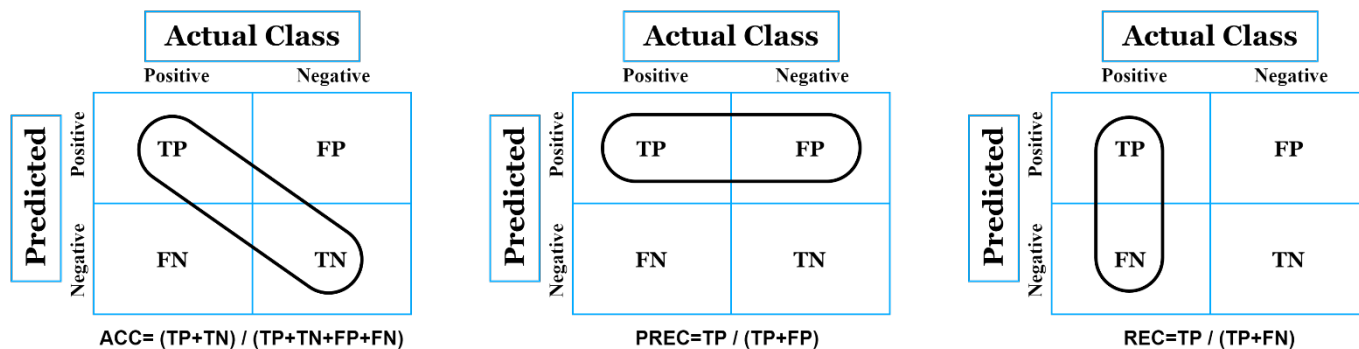


Fig. 4. Performance Metrics.

III. RESULTS AND DISCUSSION

In this section, the results of the evaluation of the various classifiers that were used in this study are presented. The statistics and results of the ontological model are also shown in Tables II, III, and Fig. 5 illustrates the performance metrics of the ontology model.

The results of this study provide a visual representation of the various metrics that are used in this research, such as precision, F-measure, Recall, and Accuracy, as shown in Fig. 6-9. Table IV also shows the results of the various classifiers that were used in this research.

TABLE II. 10-FOLD CROSS-VALIDATION FOR ONTOLOGICAL MODEL

Confusion matrix		Actual class	
		positive	Negative
Predicted class	positive	TP : 207	FP : 5
	negative	FN : 9	TN : 228

TABLE III. 50% SPLIT MODE FOR ONTOLOGICAL MODEL

Confusion matrix		Actual class	
		positive	negative
Predicted class	positive	TP : 98	FP : 6
	negative	FN : 3	TN : 117

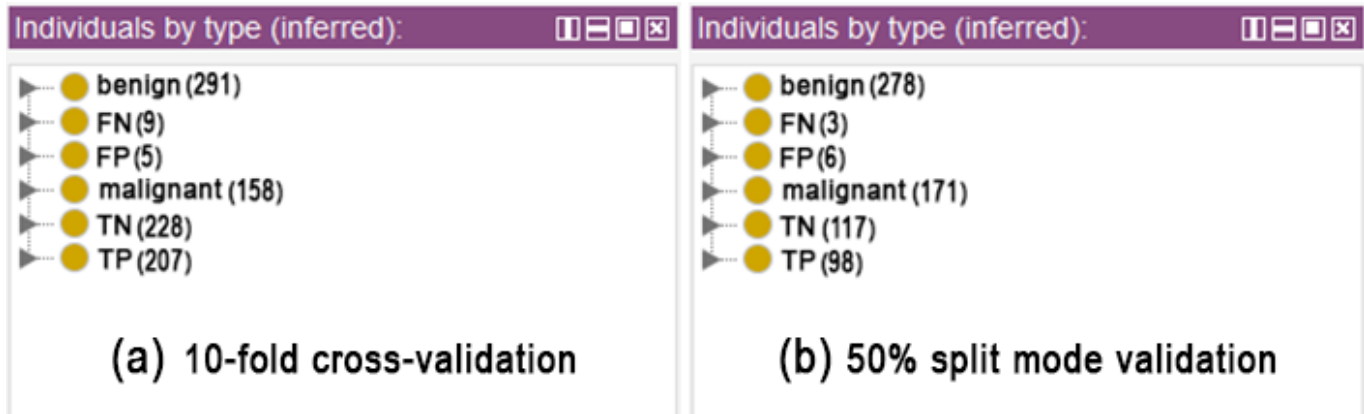


Fig. 5. Results of Inferred Concepts.

Accuracy:

The ontological model achieved the maximum value of 96.88% and Random Forest with rate of 96.00%, and 95.30 % for both Support Vector Machine and Logistic Regression in terms of 10-fold cross-validation, according to Fig. 6 and Table IV.

Precision:

The ontology classifier has the highest Precision of 97.64% in terms of 10-fold cross-validation mode, followed by Random Forest and Naïve Bayes. Concerning split test mode, the highest Precision value of 97.00% goes for ANN. More details are shown in Table IV and Fig. 7.

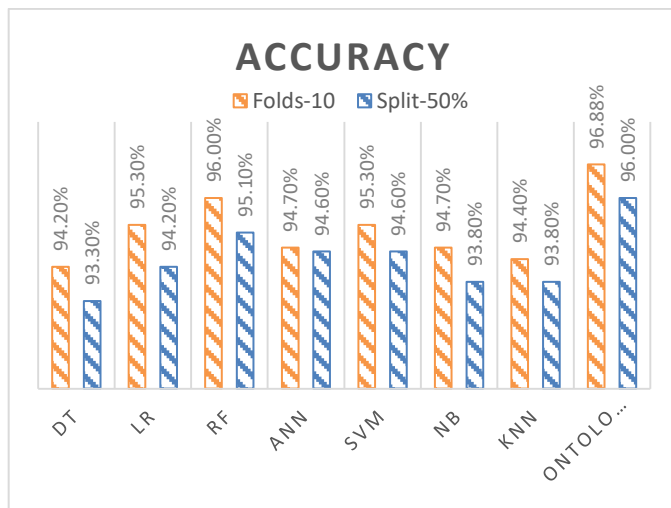


Fig. 6. Comparison Results of Accuracy.

Almost the same results using split test mode, we obtained 96.00%, 95.10% for Ontology and Random Forest consecutively, and 94.60% for both Support Vector Machine and Artificial Neural Network.

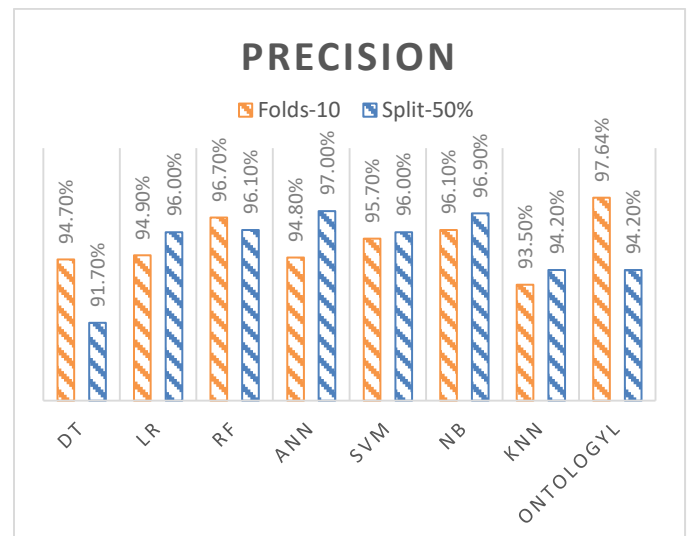


Fig. 7. Comparison Results of Precision.

Recall:

According to Fig. 8 and Table IV, the ontological model has the highest Recall values of 95.83 % and 97.00 % for both test modes, followed by RF, LR, and KNN for 10-fold cross-validation mode, and Decision Tree and Random Forest for split test mode.

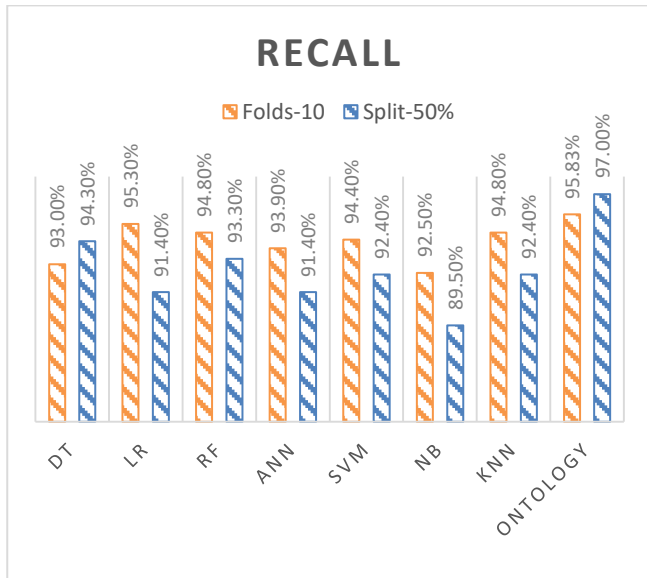


Fig. 8. Comparison Results of Recall.

F-Measure:

According to Fig. 9 and Table IV, the ontology model had the greatest value of 96 % in both test modes, followed by Random Forest in the second position and Support Vector Machine in the third position.

The experimental findings reveal that the ontology model has the highest accuracy of 96.9 %, followed by the Random Forest at 96.00 % and both Logistic Regression and Support Vector Machine at 95.30 %. In terms of the data stated above, we see no significant difference between 50%-Split and 10-Folds test modes. We conclude that the ontological model can

aid by extending the scope machine learning model. They can comprise any data kind or variation, and each diver data can be assigned to a certain job. Combining the ontological model with machine learning may provide well outcomes. The ontological model achieves results that are comparable to machine learning classifiers. Humans may interpret the findings, and the rules can be modified or added as needed. Furthermore, it supports unstructured, semi-structured, and structured data formats, allowing for more seamless data integration. It can comprise all aspects of the data modeling process, starting with schemas at the most basic level. As a result, they can handle the massive amounts of data utilized as input for machine learning training or output as outcomes. Furthermore, ontology matches any organization's aim, which might be mathematical, logical, or semantic-based. To the best of our knowledge, this is the first comparative study of the ontological model and ML in which we have integrated the ontology with ML, especially in the area of breast cancer detection. As a result, no significant comparison can be done.

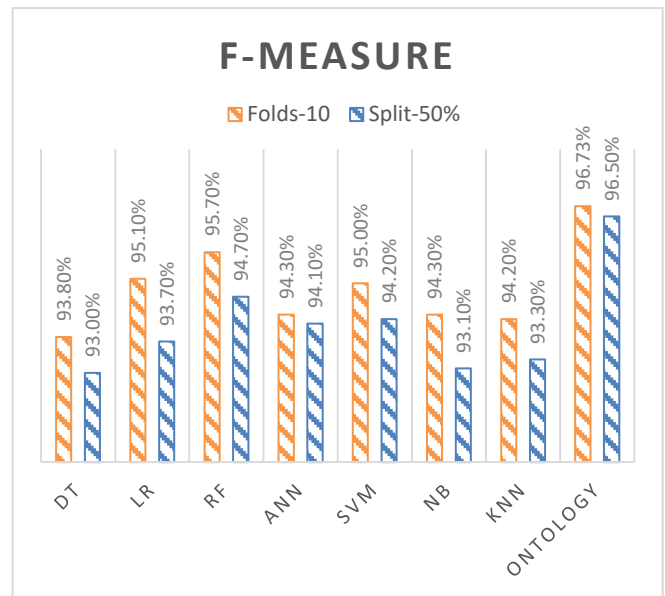


Fig. 9. F-Measure Comparison Findings.

TABLE IV. ONTOLOGICAL MODEL AND MACHINE LEARNING CLASSIFIERS RESULTS

	Accuracy		Precision		Recall		F-Measure	
	Folds-10	Split-50%	Folds-10	Split-50%	Folds-10	Split-50%	Folds-10	Split-50%
DT	0.942	0.933	0.947	0.917	0.93	0.943	0.938	0.93
LR	0.953	0.942	0.949	0.96	0.953	0.914	0.951	0.937
RF	0.96	0.951	0.967	0.961	0.948	0.933	0.957	0.947
ANN	0.947	0.946	0.948	0.97	0.939	0.914	0.943	0.941
SVM	0.953	0.946	0.957	0.96	0.944	0.924	0.95	0.942
NB	0.947	0.938	0.961	0.969	0.925	0.895	0.943	0.931
KNN	0.944	0.938	0.935	0.942	0.948	0.924	0.942	0.933
Ontology	0.969	0.960	0.976	0.942	0.958	0.970	0.967	0.965

IV. CONCLUSION

ML methods are widely employed in all scientific disciplines and have revolutionized industries all over the world. The use of machine learning techniques and algorithms in healthcare has recently advanced significantly. These approaches have shown success and may be valuable in the treatment of enduring diseases such as breast cancer. Furthermore, the Semantic Web has proven its usefulness and effectiveness in a multitude of areas, including health. As a Semantic Web component, ontology has the capability to treat concepts and relationships in the same way that humans view connected concepts.

In this research, we provided seven machine learning algorithms and an ontology model, as well as a comparison of their performance. Furthermore, two test modes are employed: 10-fold cross validation and percentage split, and several performance measures such as Accuracy, F-Measure, Precision, and Recall are employed to assess the outcomes. The findings show that the ontological model has the uppermost accuracy even when no feature selection is used. This brings us to a new search area, to which we advise and urge academics to participate and produce new insights in the same context, in order to provide additional outcomes and analysis, in order to make a forecast, recommendation, or decision, and so on. In future work, we want to improve this comparison analysis by adopting new ways to incorporate ML rules with the ontological model method, as well as regression machine learning algorithms.

REFERENCES

- [1] L. Rubinger, A. Gazendam, S. Ekhtiari, and M. Bhandari, "Machine learning and artificial intelligence in research and healthcare ☆, ☆, ☆," *Injury*, Feb. 2022, doi: 10.1016/j.injury.2022.01.046.
- [2] H. Abdel-Jaber, D. Devassy, A. Al Salam, L. Hidaytallah, and M. EL-Amir, "A Review of Deep Learning Algorithms and Their Applications in Healthcare," *Algorithms*, vol. 15, no. 2, Art. no. 2, Feb. 2022, doi: 10.3390/a15020071.
- [3] H. El Massari, S. Mhammedi, N. Gherabi, and M. Nasri, "Virtual OBDA Mechanism Ontop for Answering SPARQL Queries Over Couchbase," in *Advanced Technologies for Humanity*, Cham, 2022, pp. 193–205. doi: 10.1007/978-3-030-94188-8_19.
- [4] N. Al-Azzam and I. Shatnawi, "Comparing supervised and semi-supervised Machine Learning Models on Diagnosing Breast Cancer," *Ann. Med. Surg.*, vol. 62, pp. 53–64, Feb. 2021, doi: 10.1016/j.amsu.2020.12.043.
- [5] K. Rajendran, M. Jayabalan, and V. Thiruchelvam, "Predicting Breast Cancer via Supervised Machine Learning Methods on Class Imbalanced Data," *Int. J. Adv. Comput. Sci. Appl. IJACSA*, vol. 11, no. 8, Art. no. 8, 31 2020, doi: 10.14569/IJACSA.2020.0110808.
- [6] M. Alshammari and M. Mezher, "A Comparative Analysis of Data Mining Techniques on Breast Cancer Diagnosis Data using WEKA Toolbox," *Int. J. Adv. Comput. Sci. Appl. IJACSA*, vol. 11, no. 8, Art. no. 8, 31 2020, doi: 10.14569/IJACSA.2020.0110829.
- [7] Z. Rustam, Y. Amalia, S. Hartini, and G. S. Saragih, "Linear discriminant analysis and support vector machines for classifying breast cancer," *IAES Int. J. Artif. Intell. IJ-AI*, vol. 10, no. 1, Art. no. 1, Mar. 2021, doi: 10.11591/ijai.v10.i1.pp253-256.
- [8] W. N. L. W. H. Ibeni, M. Z. M. Salikon, A. Mustapha, S. A. Daud, and M. N. M. Salleh, "Comparative analysis on bayesian classification for breast cancer problem," *Bull. Electr. Eng. Inform.*, vol. 8, no. 4, Art. no. 4, Dec. 2019, doi: 10.11591/eei.v8i4.1628.
- [9] F. S. Khan, M. I. Abbasi, M. Khurram, M. N. H. Mohd, and M. D. Khan, "Breast cancer histological images nuclei segmentation and optimized classification with deep learning," *Int. J. Electr. Comput. Eng. IJECE*, vol. 12, no. 4, Art. no. 4, Aug. 2022, doi: 10.11591/ijece.v12i4.pp4099-4110.
- [10] N. F. Idris, M. A. Ismail, M. S. Mohamad, S. Kasim, Z. Zakaria, and T. Sutikno, "Breast cancer disease classification using fuzzy-ID3 algorithm based on association function," *IAES Int. J. Artif. Intell. IJ-AI*, vol. 11, no. 2, Art. no. 2, Jun. 2022, doi: 10.11591/ijai.v11.i2.pp448-461.
- [11] T. S. Lim, K. G. Tay, A. Huong, and X. Y. Lim, "Breast cancer diagnosis system using hybrid support vector machine-artificial neural network," *Int. J. Electr. Comput. Eng. IJECE*, vol. 11, no. 4, Art. no. 4, Aug. 2021, doi: 10.11591/ijece.v11i4.pp3059-3069.
- [12] R. A. I. Alhayali, M. A. Ahmed, Y. M. Mohialden, and A. H. Ali, "Efficient method for breast cancer classification based on ensemble hoeffding tree and naïve Bayes," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 18, no. 2, Art. no. 2, May 2020, doi: 10.11591/ijeecs.v18.i2.pp1074-1080.
- [13] S. Ara, A. Das, and A. Dey, "Malignant and Benign Breast Cancer Classification using Machine Learning Algorithms," in *2021 International Conference on Artificial Intelligence (ICAI)*, Apr. 2021, pp. 97–101. doi: 10.1109/ICAI52203.2021.9445249.
- [14] A. Atrey, N. Narayan, S. Vijh, and S. Kumar, "Analysis of Breast Cancer using Machine Learning Methods," in *2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Jan. 2022, pp. 258–261. doi: 10.1109/Confluence52989.2022.9734184.
- [15] S. Jain and P. Kumar, "Prediction of Breast Cancer Using Machine Learning," *Recent Adv. Comput. Sci. Commun.*, vol. 13, no. 5, pp. 901–908, doi: 10.2174/2213275912666190617160834.
- [16] P. P. Sengar, M. J. Gaikwad, and A. S. Nagdive, "Comparative Study of Machine Learning Algorithms for Breast Cancer Prediction," in *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Aug. 2020, pp. 796–801. doi: 10.1109/ICSSIT48917.2020.9214267.
- [17] O. Tarawneh, M. Otair, M. Husni, H. Y. Abuaddous, M. Tarawneh, and M. A. Almomani, "Breast Cancer Classification using Decision Tree Algorithms," *Int. J. Adv. Comput. Sci. Appl. IJACSA*, vol. 13, no. 4, Art. no. 4, 30 2022, doi: 10.14569/IJACSA.2022.0130478.
- [18] E. Sugiharti, R. Arifudin, D. T. Wiyanti, and A. B. Susilo, "Integration of convolutional neural network and extreme gradient boosting for breast cancer detection," *Bull. Electr. Eng. Inform.*, vol. 11, no. 2, Art. no. 2, Apr. 2022, doi: 10.11591/eei.v11i2.3562.
- [19] T. A. Assegie, R. L. Tulasi, V. Elanangai, and N. K. Kumar, "Exploring the performance of feature selection method using breast cancer dataset," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 25, no. 1, Art. no. 1, Jan. 2022, doi: 10.11591/ijeecs.v25.i1.pp232-237.
- [20] Z. Tasnim et al., "Classification of Breast Cancer Cell Images using Multiple Convolution Neural Network Architectures," *Int. J. Adv. Comput. Sci. Appl. IJACSA*, vol. 12, no. 9, Art. no. 9, Sep. 2021, doi: 10.14569/IJACSA.2021.0120934.
- [21] C. Kaul and N. Sharma, "High Accuracy Predictive Model on Breast Cancer Using Ensemble Approach of Supervised Machine Learning Algorithms," in *2021 International Conference on Computational Performance Evaluation (ComPE)*, Dec. 2021, pp. 071–076. doi: 10.1109/ComPE53109.2021.9752254.
- [22] Md. M. Islam, Md. R. Haque, H. Iqbal, Md. M. Hasan, M. Hasan, and M. N. Kabir, "Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques," *SN Comput. Sci.*, vol. 1, no. 5, p. 290, Sep. 2020, doi: 10.1007/s42979-020-00305-w.
- [23] H. EL Massari, S. Mhammedi, Z. Sabouri, and N. Gherabi, "Ontology-Based Machine Learning to Predict Diabetes Patients," in *Advances in Information, Communication and Cybersecurity*, Cham, 2022, pp. 437–445. doi: 10.1007/978-3-030-91738-8_40.
- [24] H. El Massari, Z. Sabouri, S. Mhammedi, and N. Gherabi, "Diabetes Prediction Using Machine Learning Algorithms and Ontology," *J. ICT Stand.*, pp. 319–338, May 2022, doi: 10.13052/jicts2245-800X.10212.
- [25] S. Srivastava, "Weka: A Tool for Data preprocessing, Classification, Ensemble, Clustering and Association Rule Mining," *Int. J. Comput. Appl.*, vol. 88, no. 10, pp. 26–29, Feb. 2014.

- [26] H. El Massari, N. Gherabi, S. Mhammedi, Z. Sabouri, and H. Ghandi, "ONTOLOGY-BASED DECISION TREE MODEL FOR PREDICTION OF CARDIOVASCULAR DISEASE," *Indian J. Comput. Sci. Eng.*, vol. 13, no. 3, pp. 851–859, Jun. 2022, doi: 10.21817/indjce/2022/v13i3/221303143.
- [27] A. Khamparia and B. Pandey, "Comprehensive analysis of semantic web reasoners and tools: a survey," *Educ. Inf. Technol.*, vol. 22, no. 6, pp. 3121–3145, Nov. 2017, doi: 10.1007/s10639-017-9574-5.